Impact or Responsibility?
Giving Behaviour in a Televised Natural Experiment

Inka Eberhardt Hiabu[1,2], Paul Smeets[3],
Martijn J. van den Assem[4], and Dennie van Dolder[4]

[1] Pension Research Centre (PeRCent), Copenhagen Business School
[2] Centre for Population Ageing Research (CEPAR), University of New South Wales
[3] Amsterdam Business School, University of Amsterdam
[4] School of Business and Economics, Vrije Universiteit (VU) Amsterdam
[5] Department of Economics, University of Essex

Abstract: We directly compare the influences of impact and responsibility considerations on giving behaviour. In moral philosophy, utilitarianism emphasizes the importance of the former, whereas theories of equity and desert argue for the importance of the latter. Our data are from a television show where an audience of one hundred people divides ten thousand euros among three candidates who face financial difficulties, and from independent raters who evaluated attributes of the candidates and their predicaments. We find that the well-being benefit of donations ("impact") outweighs the degree to which the candidate had control over the cause of their situation ("responsibility"). Giving increases more with impact than it decreases with responsibility, and the contribution of impact to the explanatory power of our regression models is approximately twice that of responsibility. Additionally, our analysis shows no evidence of discrimination based on age, gender, or physical attractiveness.

JEL: D91, D63, D64
Keywords: giving, impact, utilitarianism, equity, desert, responsibility, personal control
Version: February 2026

Email addresses: ieh.eco@cbs.dk, p.m.a.smeets@uva.nl, m.j.vanden.assem@vu.nl, d.vandolder@essex.ac.uk. We thank Talpa Producties B.V. for completing our collection of episodes of *Geld Maakt Gelukkig*, and Monika Riecken and Alexander Ganster for excellent research assistance. The paper has benefited from discussions with members of the BEELab in Maastricht, seminar participants at King's College London, IÉSEG School of Management Lille, University College Dublin, the University of Nottingham, Peking University Guanghua School of Management, LMU Munich, the University of Exeter, and the University of Osaka, and participants of IAREP-SABE 2023 Nice, RBFC 2024 Amsterdam, the Newcastle Experimental Economics Workshop 2024, the IAER Microeconomics Workshop 2025 Dalian, the Annual BEE Conference 2025 Sheffield, and SPUDM 2025 Lucca. Ethics approval was obtained through the BEELab (reference #18008). We gratefully acknowledge support from the Dutch Research Council (NWO; grants 452-16-011 and VI.Vidi.221E.011).

1. Introduction

Giving is a fundamental aspect of human behaviour. People donate to charitable causes, engage in volunteer work, and help others in times of need. Unpacking the determinants of giving behaviour contributes to the understanding of decision-making processes and social preferences, and can inform policies aimed at enhancing economic and social welfare.

There are different normative views on giving. According to utilitarianism, a consequentialist strand of moral philosophy advanced by Jeremy Bentham and John Stuart Mill, individuals should give to maximize societal well-being. This perspective assumes that well-being can be quantified and aggregated across individuals, and provides the philosophical foundation for welfare economics. Utilitarian thinking is central to the effective altruism movement, which advocates directing donations to charities that generate the highest impact per dollar (Singer, 2009; MacAskill, 2015).

The empirical evidence on the descriptive validity of utilitarianism in giving behaviour is mixed. On the one hand, there is research showing that donation decisions are influenced by quality ratings of charities (Gordon et al., 2009; Yörük, 2016; Adena et al., 2019; Caviola et al., 2020), and that highlighting the impact increases the emotional rewards of giving (Aknin et al., 2013). On the other hand, there is also evidence that people tend to naively focus on avoiding overhead costs rather than maximizing impact (Baron and Szymanska, 2011; Caviola et al., 2014; Gneezy et al., 2014), that people's willingness to pay for information about the impact of donations is low (Null, 2011; Metzger and Günther, 2019), and that donations are often only weakly influenced by the impact they have (Hsee and Rottenstreich, 2004; Null, 2011; Hasford et al., 2015; Karlan and Wood, 2017; Berman et al., 2018).

Whereas utilitarianism exclusively considers the outcome of actions, theories of equity and desert put emphasis on how these outcomes are achieved. Rooted in Aristotle's theory of distributive justice and John Locke's theory of natural law/desert, these theories argue that the fairness of outcomes should be evaluated on the basis of individuals' responsibility for the outcomes (Konow, 2001, 2003).[1] Since giving can be viewed as a form of redistribution of outcomes, this fairness principle implies that people should be more inclined to give when a recipient's need results from bad luck rather than from factors within their control.

There is strong empirical evidence for the importance of equity and desert considerations in giving behaviour. For instance, people who believe that the poor are responsible for their own situation are less supportive of redistribution than those who view poverty as resulting from circumstances beyond

[1] In economics, this moral view is referred to by various names, including the accountability principle (Konow, 1996, 2000), responsibility-sensitive egalitarianism (Schokkaert and Devooght, 2003; Robson et al., 2025), luck egalitarianism (Mollerstrom et al., 2015; Roemer and Trannoy, 2016), and liberal egalitarianism (Cappelen et al., 2007; Cappelen and Tungodden, 2009).

individual control (Fong, 2001; Alesina and La Ferrara, 2005; Alesina and Giuliano, 2011; Cohn et al., 2023). Furthermore, controlled experiments show that people are more likely to redistribute money to someone who experienced a relatively poor outcome due to chance than to someone who is personally responsible for it (Konow, 2000; Cappelen et al., 2007, 2013; Trhal and Radermacher, 2009; Cappelen et al., 2010; Krawczyk, 2010; Almås et al., 2020). Similarly, experimental studies on charitable giving that manipulate the information about recipients find that donors are more willing to give when difficulties appear to stem from circumstances beyond recipients' control, such as disability, rather than from recipients' own choices (Fong, 2007; Fong and Oberholzer-Gee, 2011; Gangadharan et al., 2023). Last, people are more inclined to donate to research on hereditary cancer than to research on lifestyle-related cancer (Safra et al., 2019; Melkonyan et al., 2021).

In social psychology, equity and desert considerations are often studied through the lens of attribution theory, which focuses on how people perceive the causes of events and how these perceptions influence behaviour (Heider, 1958; Jones et al., 1972; Weiner, 1974, 1985, 1986). The key causal property (or "attribution dimension") hypothesized to affect helping behaviour is the degree of controllability: when the cause of a negative outcome is within a person's control, they are seen as responsible for it and less deserving of help. Consistent with this view, many studies empirically document that people's judgments of a recipient's personal control over the cause of their situation determine their willingness to give (Weiner, 1980a, 1980b, 1995, 2006; Reisenzein, 1986; Schmidt and Weiner, 1988; Betancourt, 1990; Greitemeyer and Rudolph, 2003; Rudolph et al., 2004; Osborne and Weiner, 2015; Tscharaktschiew and Rudolph, 2016).

Altogether, the existing literature thus provides mixed evidence on the importance of utilitarian (or "impact") considerations in giving behaviour, and strong evidence for the importance of equity and desert (or "responsibility") considerations, suggesting that people place greater weight on the latter when deciding whom to help. However, to the best of our knowledge, no empirical research has directly compared the significance of these two philosophical perspectives on giving behaviour.

The present paper uses the Dutch TV show *Geld Maakt Gelukkig* ("Money Buys Happiness"; henceforth: GMG) to examine and compare how perceptions of impact and responsibility influence donation decisions in a quasi-controlled field setting with stakes that are consequential for the recipients. In GMG, three individuals (henceforth: candidates) appeal to a studio audience of one hundred people for financial help to address a difficulty they face. Each audience member is endowed with €100 and must donate this amount to one of the three candidates at the end of the episode. Each candidate thus receives a share of the total of €10,000.

To measure people's perceptions of the well-being impact of donations and candidates' responsibility for their situation, we hired independent raters. These raters watched excerpts from the show in which

the candidates and their financial difficulties are presented. For impact, we then asked them how well-being would be affected by a gift of €100 (the donation of a single audience member) and by a gift of €10,000 (the total amount to be divided) to a given contestant. To measure responsibility, we employed the Revised Causal Dimension Scale (CDSII; McAuley et al., 1992). Personal control is one of the four dimensions of this scale, and captures the extent to which the cause is controllable by the individual of interest. Raters also evaluated the physical attractiveness and age of the candidates.

To compare how the giving behaviour of the audience is influenced by impact and responsibility, we first regress the donation outcomes on the different ratings. We account for the specific structure of our data by employing the Dirichlet-multinomial model (Shonkwiler and Hanley, 2003; Guimarães and Lindrooth, 2007), and find that donations increase strongly with impact and decrease moderately with responsibility. We then use dominance analysis to assess the relative importance of impact and responsibility considerations, which shows that impact is approximately twice as important as responsibility. These results hold across a wide range of analytical choices and are further supported by a replication study. Contrary to the picture emerging from the existing literature, our empirical comparison thus indicates that utilitarianism provides a stronger descriptive account than theories of equity and desert. Additionally, we find no evidence of discrimination based on age, gender, or physical attractiveness.

GMG provides a unique opportunity to study giving behaviour in a well-defined choice task that affects the well-being of real recipients facing real financial difficulties. The reasons candidates request money and the underlying causes are highly diverse, offering a rich variety of cases for studying the roles of impact and responsibility. The stakes are consequential for the candidates, and far exceed those normally used in conventional experiments.[2] Audience members make their donation decisions under controlled conditions: they always choose among three recipients whose cases are presented in a uniform format, and they always donate €100 individually and €10,000 in total. Their choices are quasi-anonymous, they have no personal connection with the candidates, and they have no information about the candidates beyond what is provided in the show.[3] The audience is visibly diverse in terms of gender, age, and ethnicity, and our results therefore likely reflect the behaviour of a broad (middle-class) cross-section of the general population. A limitation is that we cannot link donations to individual members, which makes it impossible to study heterogeneity in giving behaviour. An important feature of GMG is that the donors are endowed with the money they donate, and cannot keep it for

[2] If we had run this giving experiment ourselves, the total cost of donations to candidates alone would have been more than half a million euros.

[3] Choices are not fully anonymous because candidates can see the audience and attentive TV viewers can catch glimpses of the audience, but beyond the overall distribution of donations it remains unknown who gave to whom.

themselves. Combined with the absence of any personal connection with the candidates, this ensures that personal interests do not influence decisions. As a result, our study aligns with the established approach in laboratory experiments of examining moral preferences by observing the decisions of impartial spectators rather than those of implicated stakeholders (Konow, 2000, 2005, 2009; Dickinson and Tiefenthaler, 2002; Cappelen et al., 2013; Cettolin and Riedl, 2017; Almås et al., 2020; Cappelen et al., 2022).

Our work connects to a broader literature that uses TV shows to investigate decision making. This approach has been used to study, for example, decision making under risk (Gertner, 1993; Metrick, 1995; Post et al., 2008; Jetter and Walker, 2018), strategic decision making (Bennett and Hickman, 1993; Berk et al., 1996; Klein Teeselink et al., 2024), discrimination (Levitt, 2004; Antonovics et al., 2005; Belot et al., 2012), cooperation (List, 2004, 2006; Belot et al., 2010; van den Assem et al., 2012; Turmunkh et al., 2019), bargaining (van Dolder et al., 2015), and willingness to compete (Hogarth et al., 2012; Buser et al., 2023). Related to our study, Gurevich and Kliger (2013) study cooperative behaviour in the Israeli TV game show "The Manipulation", in which contestants who claim to be in financial need play a variant of the prisoner's dilemma. They focus on attribution theory and find that cooperation decreases with perceptions of controllability.

The paper proceeds as follows: Section 2 describes the TV show and our data, Section 3 presents the main analyses and results, Section 4 considers the robustness of our results, and Section 5 concludes with a summary and a discussion of our findings.

2. TV Show and Data

2.1 TV show

GMG was developed and produced by the Dutch media company Talpa, and broadcast on weekdays in 2014 and 2015 on the Dutch TV channel SBS6. The show comprised two seasons, for a total of 55 episodes.[4] We downloaded most episodes from the internet and obtained the remaining ones from the production company.

In each episode, three candidates appeal to the studio audience for money to resolve a financial difficulty. The audience consists of one hundred people, and is visibly diverse in terms of gender, age,

[4] The first season, consisting of 20 episodes, ran from 7 July until 1 August 2014. The second season, consisting of 35 episodes, was televised between 5 January and 6 February 2015 (25 episodes) and between 6 July and 17 July 2015 (10 episodes). There are no material differences between the two seasons.

and ethnicity.[5] Each audience member is endowed with €100 and must donate this amount to one of the three candidates at the end of the episode. In total, €10,000 is divided among the three candidates.

Episodes are just over twenty minutes long and start with presentations of the three cases. The three candidates are introduced one by one, along with the problem they face and the amount of money they are requesting. This introduction is delivered through the combination of a short video documentary and an unstructured panel interview. The video, shot on location by the show's producer and lasting about two-and-a-half minutes, features the candidate and a voice-over explaining why financial help is needed and how much is being requested. The subsequent panel interview takes place in the studio and features two semi-famous experts—a lawyer and a financial coach—briefly questioning the candidate about the situation.[6]

After the problems of the candidates have been outlined and explained, all audience members simultaneously select the candidate they wish to support, using their personal voting device.[7] Episodes conclude with a brief discussion of how much of the total of €10,000 each candidate received.

The total pool of 165 cases is highly diverse and includes, for example, (i) a terminally ill and disabled person who wants to visit her family's country of origin to find a place for her ashes to be scattered, (ii) a young adult with a troubled childhood and accumulated debts who is on the verge of homelessness, (iii) an unemployed single mother of two in a wheelchair who is dependent on a worn-out modified car, and (iv) a family man who is unable to meet his mortgage obligations due to irresponsible behaviour of a former business partner. Appendix A discusses three other example cases in detail. Typically, a candidate or their family is the primary beneficiary of the money awarded, and the case is represented on the stage by a single person. In 12 instances, however, the candidate represents an outside beneficiary—usually a small charitable foundation closely connected to them. In 20 instances, two people appear on the stage for a single case, yet only one participates in the interview with the panellists; our analyses focus on this primary candidate.[8]

[5] According to a call for audience members, the producer required that audience members be between 18 and 55 years of age. They were provided with complimentary drinks and snacks, and a camera was raffled off.

[6] Of course, the information in the introductions is likely incomplete and may even present a distorted picture. This does not pose a problem for our study, as we examine how donors act on their perceptions. Note that in the real world, donors are often similarly faced with uncertainty regarding the accuracy and completeness of the available information.

[7] Audience members sit close together, without dividers. We therefore cannot rule out the possibility of interaction between them. However, they are expected to remain silent during the recordings, and asked to make their donation decisions individually.

[8] Robustness checks in Subsection 4.2 show that our results do not materially change when we exclude those 12 or 20 cases.

2.2 Rating sessions

We organized a series of rating sessions to obtain case-specific measures of (i) the overall well-being impact of a donation, (ii) the candidate's responsibility for their situation, (iii) their physical attractiveness, and (iv) their age. We recruited 80 Dutch-speaking raters from the student population of Maastricht University in the Netherlands. Each rater participated in one of twelve sessions held at the Behavioral and Experimental Economics Laboratory (BEELab) at Maastricht University.

At the start of each session, we handed out the instructions (see Appendix B), read them aloud, and asked whether anything was unclear. No questions were asked in any of the sessions. The raters then drew cards to randomly determine their individual cubicle, after which they entered their cubicle and started the pre-installed questionnaire on the computer.

Each rater was assigned nine different cases, corresponding to the complete set of cases of three different episodes. They first viewed and assessed the three cases from one episode, then those from a second, and finally those from a third. We selected and ordered the three episodes semi-randomly, ensuring that every episode from the full sample appeared at least once in each of the three positions, and that it was assessed by at least four and no more than five different raters. For each episode, we preserved the original order of the three cases to avoid inconsistencies between the presentation sequences shown to the raters and those seen by the audience.

For each case, the raters first watched the segment from the episode that included both the documentary and the subsequent panel interview. The average length of these video clips was 5 minutes and 20 seconds (shortest: 4:02, longest: 6:47). Raters could proceed to the next screen only after the video had played for at least four minutes, and were thus effectively required to watch every video in full. They were shown only these excerpts and were not given any description of the show.

After watching the video of a case, the raters were asked to assess the well-being impact of a hypothetical gift of a given size to the candidate, and to answer various questions about the cause of the candidate's situation (explained below). They were also asked to estimate the candidate's age in years, and to rate the candidate's physical appearance on a scale from 1 to 10 (where 1 indicated "extremely unattractive" and 10 "model beautiful or handsome").

In the final part of the questionnaire, we asked the raters for their age, gender, faculty of study, and familiarity with the show. At the end of the session, we paid everyone a show-up fee of €13. After all sessions were completed, we also raffled €100 among the entire pool of raters.

The average age of the raters was 20 years (min: 18, max: 28). Forty-four percent were female and 56 percent were male; no one identified as other. Eighty-five percent were enrolled at the School of

Business and Economics. Most raters indicated that they had never seen the show (82%), a minority had seen one or more episodes (14%), and a few were not sure (4%).[9]

2.3 Impact

We asked two questions about the perceived impact of a gift on the candidate's well-being: one for a gift of €100 and one for a gift of €10,000. A priori, it was unclear which of the two amounts would be most relevant. €100 is what each audience member individually decides upon, and thus the amount that determines the difference their choice can make. €10,000 is the maximum the candidate could receive from all audience members combined, and the amount that would typically fully cover their financial need. The questions were asked as follows:

"In general, gifts increase the well-being of the recipient and people who are indirectly involved. Assume that this impact can be measured, and that the overall impact of a gift is the sum of the increases in well-being for all people involved.

Imagine a group of 100 people, consisting of the candidate and 99 people who are randomly drawn from the Dutch population. Assume that people can be ranked in terms of the overall impact from receiving a monetary gift.

What would the rank of the candidate be if those 100 people are ordered from lowest overall impact (1) to highest overall impact (100) of a gift of €100?

What would the rank of the candidate be if the 100 people are ordered from lowest overall impact (1) to highest overall impact (100) of a gift of €10,000? (Note that this rank can be lower than, higher than, or equal to your previous answer where the gift was €100.)"

The wording of the two questions aligns with utilitarian principles: well-being effects can be quantified and aggregated, and what matters is the overall well-being effect across all people affected. The percentile scale with the Dutch general population as the reference group served to promote the comparability of responses between raters. In the first question we underlined the amount of €100, in the second we underlined €10,000. Raters reported the two percentiles using sliders.

[9] These statistics describe 79 of the 80 raters. For one rater the information about their age, gender, study, and familiarity with the show is missing in our data.

Table 1: Revised Causal Dimension Scale (CDSII)

The table shows the CDSII section of the questionnaire. Raters reported on the nine-point Likert scale by selecting the respective radio button. The scores for personal control are obtained by aggregating items 2, 4 and 10 (external control: 5, 8 and 12; internal locus of causality: 1, 6 and 9; stability: 3, 7 and 11; item numbers were not shown in the actual questionnaire). Adapted with minor modifications from McAuley et al. (1992).

Think about the cause or causes you have just described. The items below concern your impressions or opinions of this cause or these causes.										
Is the cause, or are the causes, something…										
		9	8	7	6	5	4	3	2	1	
1	that reflects an aspect of the candidate	○	○	○	○	○	○	○	○	○	that reflects an aspect of the situation
2	manageable by the candidate	○	○	○	○	○	○	○	○	○	not manageable by the candidate
3	permanent	○	○	○	○	○	○	○	○	○	temporary
4	the candidate can regulate	○	○	○	○	○	○	○	○	○	the candidate cannot regulate
5	over which others have control	○	○	○	○	○	○	○	○	○	over which others have no control
6	inside of the candidate	○	○	○	○	○	○	○	○	○	outside of the candidate
7	stable over time	○	○	○	○	○	○	○	○	○	variable over time
8	under the power of other people	○	○	○	○	○	○	○	○	○	not under the power of other people
9	about the candidate	○	○	○	○	○	○	○	○	○	about others
10	over which the candidate has power	○	○	○	○	○	○	○	○	○	over which the candidate has no power
11	unchangeable	○	○	○	○	○	○	○	○	○	changeable
12	other people can regulate	○	○	○	○	○	○	○	○	○	other people cannot regulate

2.4 Responsibility

For the assessment of the candidate's responsibility for their predicaments we used the personal control dimension of the Revised Causal Dimension Scale (CDSII; McAuley et al., 1992). The CDSII serves to measure the causal dimensions of attribution theory (Weiner, 1985, 1986), an influential theory in social psychology on how people perceive the causes of events and how these perceptions influence behaviour. It distinguishes four dimensions of causal attributions.[10] Personal control captures the extent to which the cause is controllable by the individual in question. The other three dimensions are external control (controllability by others), internal locus of causality (extent to which the cause is internal rather than external to the individual), and stability (how constant the cause is over time). Each dimension is measured by three items. For the sake of completeness and to allow controlling for external control, internal locus, and stability in our empirical analyses, we included the full CDSII in the questionnaire. Following McAuley et al. (1992), we first asked the raters to briefly describe the cause of the candidate's situation in their own words before we presented the twelve items to them.

[10] Russell (1982) developed the original Causal Dimension Scale as a three-dimensional measure, distinguishing locus of causality, stability, and controllability. In response to concerns regarding the latter dimension, McAuley et al. (1992) introduced the four-dimensional Revised Causal Dimension Scale (CDSII), which separates controllability into personal control and external control.

Table 1 presents the exact wording for each item. The only modification from the version proposed by McAuley et al. (1992) is our designation of the central subject as "the candidate". Raters reported their causal beliefs on nine-point Likert scales by selecting the respective radio buttons. Higher values indicate that the cause is more controllable by the candidate (personal control; items 2, 4, and 10), more controllable by others (external control; 5, 8, and 12), more internal rather than external to the candidate (internal locus; 1, 6, and 9), or more stable over time (stability; 3, 7, and 11).

2.5 Summary statistics

In addition to the data obtained through the rating sessions, we hand-collected candidates' gender, race (white vs. non-white), and age, as well as the amounts they requested and received. The age of about half of the candidates (53%) was mentioned in the episode; for the others, we rely on the average of the age estimates provided by our raters. The average number of raters per case was 4.36; 105 cases were assessed by four, 60 by five. To obtain case-level summary statistics, we use the simple average of the four or five scores.

As shown in Table 2, Panel A, approximately two-thirds of the candidates are female (66%), and the vast majority are white (94%). The average age of the candidates is 38 years (min: 18, max: 71). The average rating of their physical attractiveness is 5.14, slightly below the scale midpoint of 5.5 (min: 2.25, max: 8.60). The average requested amount is €5,087 (min: €1,200, max: €10,000). Seventeen candidates (10%) asked for the maximum of €10,000 (untabulated). As the audience always divides €10,000 among three candidates, the average amount received is €3,333. The largest amount received by a candidate is €8,300, while one candidate received nothing. Just over a quarter of the candidates (28%) received the amount they requested or more (untabulated). Appendix D shows histograms of the requested and received amounts.

On average, raters perceived the candidates to be in the 38th percentile of the broader population in terms of the well-being impact of receiving €100. For €10,000, the average corresponds to the 68th percentile. The relatively low average for €100 may reflect an anticipated contrast effect: the much larger requested amount that resolves a problematic financial need likely constitutes a high-impact comparison point, and €100 seems to make little difference in that light.[11] The relatively high average impact of €10,000 makes sense, considering that this amount resolves a significant issue for the candidate.

[11] This conjecture is supported by the negative correlation between *Amount requested* and *Impact €100* (see Table C1 in Appendix C).

Table 2: Summary statistics

The table reports summary statistics for unstandardized (Panel A) and standardized (Panel B) variables describing the 165 candidates, their cases, and the video clips. In Panel B, all variables are after the standardization of (underlying item) scores at the rater level. *Age* is the candidate's age in years. *Female* is a dummy variable that takes the value of 1 if the candidate is female. *White* is a dummy variable that takes the value of 1 if the candidate is white. *Attractiveness* is the average rating of the candidate's physical appearance, assessed by raters on a scale from 1 to 10. *Amount requested* and *Amount received* are the sums of money the candidate asked for and received from the audience members, respectively, in euros. *Impact €100* and *Impact €10,000* are the average ranks of the impact of a gift of €100 and €10,000, respectively, on the candidate's well-being, assessed by raters as a percentile with the Dutch general population as the reference group. *Personal control*, *External control*, *Internal locus*, and *Stability* are the four causal attribution dimension scores for the candidate's case, each calculated as the average of the scores assigned by raters to the three underlying CDSII items (McAuley et al., 1992), on a nine-point Likert scale. *Video duration* is the length of the video clip of the candidate's case that was watched by the raters, in minutes and seconds. *Raters* is the number of raters that assessed the candidate's case.

	Mean	SD	Min	Q1	Median	Q3	Max
Panel A: Unstandardized							
Candidate characteristics							
Age	38.2	13.3	18.0	26.0	37.0	48.8	71.3
Female	0.66	0.47	0.00	0.00	1.00	1.00	1.00
White	0.94	0.24	0.00	1.00	1.00	1.00	1.00
Attractiveness	5.14	1.35	2.25	4.00	5.00	6.00	8.60
Money amounts							
Amount requested	5,087	2,295	1,200	3,500	4,900	6,000	10,000
Amount received	3,333	1,767	0	1,900	3,200	4,500	8,300
Impact							
Impact €100	38.1	13.8	8.3	27.5	37.3	47.8	82.0
Impact €10,000	67.6	14.1	28.8	59.3	68.8	76.6	97.2
Causal attributions							
Personal control	4.80	1.48	1.42	3.75	4.93	5.87	8.07
External control	4.98	1.35	1.58	4.00	4.92	5.87	8.17
Internal locus	5.19	1.60	1.75	4.08	5.08	6.50	8.75
Stability	4.65	1.41	1.67	3.50	4.50	5.73	8.25
Other							
Video duration	5:20	0:37	4:02	4:53	5:18	5:44	6:47
Raters	4.36	0.48	4.00	4.00	4.00	5.00	5.00
Panel B: Standardized							
Candidate characteristics							
Attractiveness	0.00	0.71	-1.58	-0.55	-0.06	0.52	1.71
Impact							
Impact €100	0.00	0.52	-1.13	-0.33	-0.01	0.29	1.72
Impact €10,000	0.00	0.54	-1.54	-0.36	-0.02	0.41	1.47
Causal attributions							
Personal control	0.00	0.58	-1.45	-0.41	0.06	0.37	1.29
External control	0.00	0.53	-1.33	-0.40	-0.01	0.39	1.35
Internal locus	0.00	0.57	-1.39	-0.41	-0.01	0.43	1.57
Stability	0.00	0.51	-1.08	-0.40	-0.03	0.38	1.25

The four causal attribution dimension scores are roughly centred around the midpoint of the nine-point Likert scale.[12] Both the two impact measures and the four causal attribution dimensions exhibit considerable variation across cases. Table C1, Panel A in Appendix C reports the correlation coefficients between the variables.[13]

To illustrate the diversity of the cases and to provide a clearer sense of the data, Appendix A describes three cases from one episode, along with raters' judgments of attractiveness, their evaluations of the impact of giving money, and their causal attributions.

2.6 Standardization of ratings

Raters evaluated the impact of €100 and €10,000 on percentile scales, the 12 CDSII items on nine-point Likert scales, and physical attractiveness on a ten-point Likert scale. Raters may differ in how they interpret the scales, which could hamper the comparability of responses across raters. To address this, we followed the procedure used by Landry et al. (2006, 2010) and standardized each rater's score on each item. Specifically, we calculated the standardized ratings as $\alpha_{jrs}^{ST} = (\alpha_{jrs} - \bar{\alpha}_{rs})/\sigma_{rs}$, where α_{jrs} is rater r's score for candidate j on item s, $\bar{\alpha}_{rs}$ is rater r's average score on item s across the nine candidates they saw, and σ_{rs} is the standard deviation of rater r's scores on item s. This procedure produces item scores that follow a standard normal distribution at the rater level.

To obtain our two impact measures—one for the impact of €100 and one for €10,000—we averaged the standardized scores on these items across the four or five raters who evaluated the given case. We did the same for physical attractiveness. For each of the four CDSII dimensions, we first averaged the standardized scores on the three items of the dimension within raters, and then averaged across the raters. Robustness analyses in Subsection 4.3 show that our results are not dependent on this standardization.

[12] A candidate's score on a dimension is the simple average of their scores on the three underlying items.

[13] The two impact measures are positively correlated. The CDSII's two controllability dimensions are negatively correlated with the stability of the cause. Locus of causality is assessed as more internal with higher personal control, and more external with higher external control. The correlation between personal and external control is close to zero, supporting the case for separating the controllability dimension of the original Causal Dimension Scale into these components, and thus for the CDSII (Russell, 1982; McAuley et al., 1992). The perceived well-being impact of €10,000 is higher when personal control is lower and when the locus of causality is more external. A larger requested amount is associated with a higher impact of €10,000, a lower impact of €100, and less personal control. Physical attractiveness judgments are negatively correlated with the candidate's age, and are higher for female candidates than for male candidates.

2.7 Standardized variables

Table 2, Panel B provides the summary statistics for the standardized variables, and Table C1, Panel B in Appendix C provides the corresponding correlation coefficients.[14,15]

A principal-component analysis of the standardized scores for the 12 CDSII items confirms the presence of four principal factors (Kim and Mueller, 1978). After applying varimax rotation to produce orthogonal factors (Kaiser, 1958), the factor loadings align closely with the four CDSII dimensions. Our primary variable of interest, personal control, exhibits strong internal reliability: Cronbach's alpha is 0.83. For the other three dimensions, the Cronbach's alpha values are 0.77 (external control), 0.74 (internal locus), and 0.64 (stability).

To assess inter-rater agreement, we compute intraclass correlation coefficients (ICCs) for the standardized variables (Shrout and Fleiss, 1979; McGraw and Wong, 1996).[16] For the CDSII dimensions and attractiveness, the ICCs are relatively high: 0.78 for personal control, 0.65 for external control, 0.85 for internal locus, 0.70 for stability, and 0.78 for attractiveness. For the impact measures, the inter-rater agreement is relatively weak: 0.29 for the impact of €100 and 0.38 for the impact of €10,000. These two ICCs suggest considerable measurement error, which may attenuate the estimated impact coefficients in our regression analyses (Bound et al., 2001; Hausman, 2001). Reassuringly, the replication study discussed in Subsection 4.5 reports high ICCs for both impact and personal control, and this study yields similar results.

3. Analyses and Results

3.1 Descriptive analysis

We use ternary plots to provide an initial view of the relationships between the variables of interest and the proportions of the prize pool that candidates received. Figure 1, Panel A depicts the percentage share of the prize pool received by candidates ranking highest, middle, or lowest on the impact of €100. Each point in the triangle represents one episode. Collectively, the points are tilted towards the top and bottom-left vertices, indicating that the highest- and middle-ranked candidates received

[14] Note that the standard deviations in Table 2, Panel B are smaller than one. The variables are averages at the case level, which, by construction, vary less than the underlying standardized scores at the rater level.

[15] After standardization, the correlation between the two impact variables is even stronger, and personal control is negatively correlated with both impact variables (rather than only with the impact of €10,000).

[16] To compute $\mathrm{ICC}(2,k) = \sigma^2_{case}/(\sigma^2_{case} + (\sigma^2_{residual}/k))$, we use a two-way random-effects model that decomposes the overall variance into between-case variance and residual variance, and set *k* at 4.36 (the average number of raters per case).

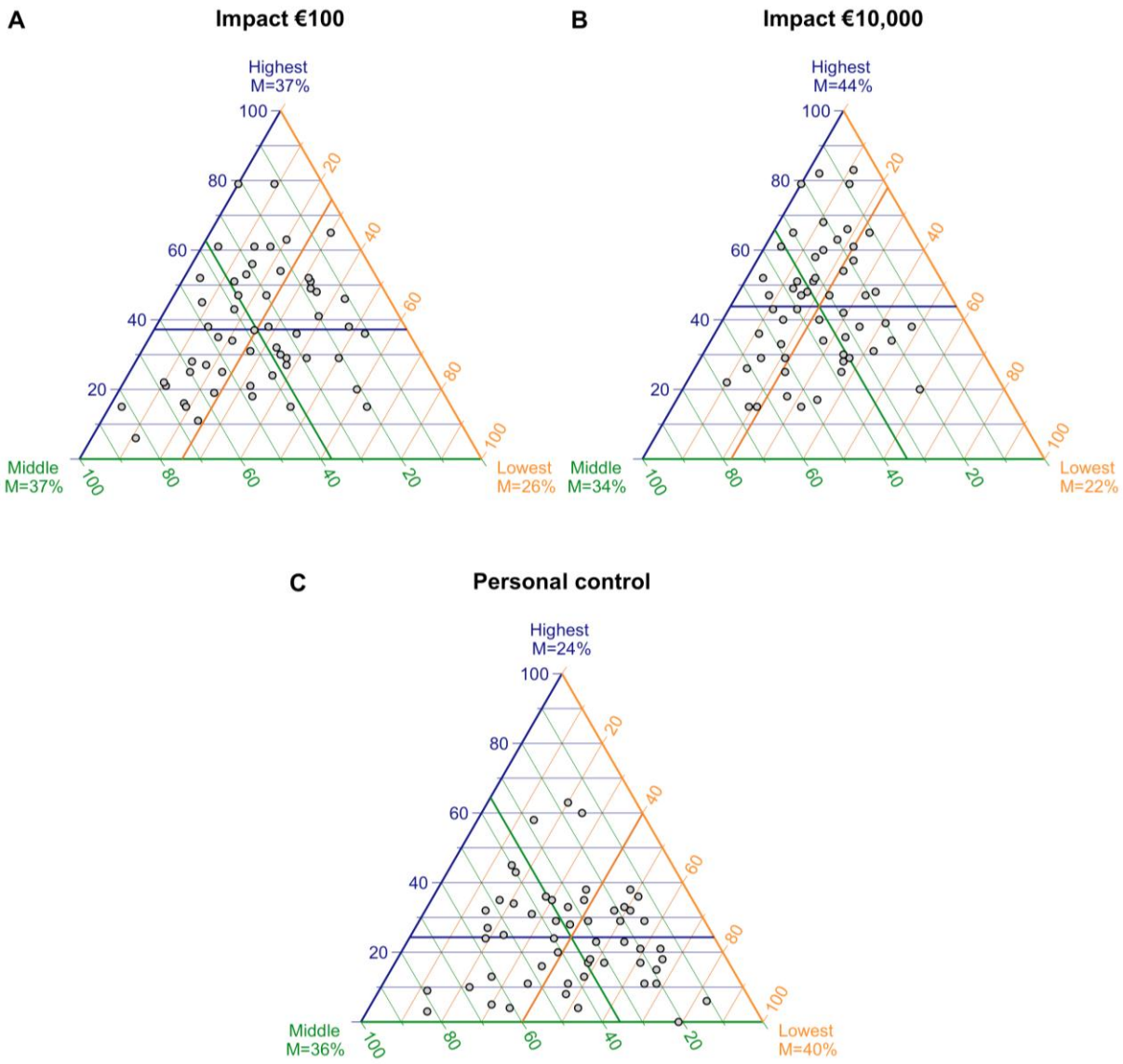

Figure 1: Ternary plots of donations in the show by rankings on *Impact €100*, *Impact €10,000,* and *Personal control*. The figure shows ternary plots for the percentage shares received by candidates ranking highest, middle, or lowest on *Impact €100* (Panel A), *Impact €10,000* (B), and *Personal control* (C). Variables are defined as in Table 2, and after the standardization of underlying item scores at the rater level.

relatively large shares. The solid lines represent the averages, showing that the highest- and middle-ranked candidates both received 37 percent, while the lowest-ranked candidate received 26 percent.

Panel B shows the donation distributions when candidates are ranked on the impact of €10,000. For this amount, the relation between rank and share seems even stronger: compared to Panel A, the points are located more closely to the top vertex and further from the bottom-right vertex. On average, the highest-ranked candidate received 44 percent, the middle-ranked 34 percent, and the lowest-ranked 22 percent.

Panel C shows the divisions when candidates are ranked on their personal control over the cause of their situation. The points are tilted towards the bottom of the triangle, particularly towards the bottom-right, indicating that more money went to those with less personal control. On average, the candidate with the most control received 24 percent, those with the intermediate degree of control 36 percent, and those with the least control 40 percent.

For completeness, Figure E1 in Appendix E shows the proportions when candidates are ranked according to the other three causal attribution dimensions (external control, internal locus, and stability), attractiveness, and age. For external control, attractiveness, and age, the shares seem unrelated to the ranking: the points are scattered around the centre of the triangle. For internal locus and stability there does appear to be a relation, with those who rank low on internal locus and those who rank high on stability receiving larger shares.[17] Additionally, a simple comparison indicates that, on average, women received somewhat larger shares than men (35.3% vs. 29.6%).

As shown in Table C1, Panel B in Appendix C, there are several strong correlations between the variables. The two impact variables are positively correlated, and both are negatively correlated with personal control. Stability is positively correlated with the impact variables and negatively with personal control, and internal locus is negatively correlated with the impact of €10,000 and positively with personal control. To disentangle the effects of the different variables on the donation decisions, we turn to multivariate regression analysis.

3.2 Modelling the giving decisions

Audience members decided individually to whom to give their €100, but the variation in our data is at the case-episode level. Within an episode, all audience members faced the same choice set, and their choices are grouped as vectors of counts indicating how many audience members donated to each of the three candidates. To analyse the data, we employ the Dirichlet-multinomial model proposed by Shonkwiler and Hanley (2003) and Guimarães and Lindrooth (2007). This model is well-suited for such grouped choice data and can be applied within a standard random utility maximization framework (McFadden, 1974). It accounts for potential correlations between the choices of audience members within an episode that are not captured by observables. Additionally, the model ensures that the probabilities that an audience member donates to the different candidates sum to one and that these are determined by the candidate's characteristics relative to those of the other candidates in the episode.

[17] When the ranking is based on internal locus, the highest-ranked candidate on average received 29 percent, the middle-ranked 33 percent, and the lowest-ranked 38 percent. For stability, these percentages are 40, 34, and 26, respectively.

Following Guimarães and Lindrooth (2007), we assume that the decision utility that audience member i in episode g attaches to donating their €100 to candidate j is given by:

$$U_{ijg} = \beta' x_{jg} + \eta_{jg} + \varepsilon_{ijg} \tag{1}$$

where β is a vector of unknown parameters, x_{jg} represents observable characteristics of the candidates that influence the audience members' giving decisions, η_{jg} is a random variable accounting for unobserved case-specific characteristics that influence all audience members, and ε_{ijg} is an independent random variable.

If ε_{ijg} is independently and identically distributed with a Type I extreme-value distribution and audience members donate to the candidate for whom U_{ijg} is highest, then the probability that an audience member in episode g donates to candidate j is given by:

$$\tilde{p}_{jg} = \frac{\exp(\beta' x_{jg} + \eta_{jg})}{\sum_{j=1}^{J} \exp(\beta' x_{jg} + \eta_{jg})} = \frac{\tilde{\lambda}_{jg} \exp(\eta_{jg})}{\sum_{j=1}^{J} \tilde{\lambda}_{jg} \exp(\eta_{jg})} \tag{2}$$

where $\tilde{\lambda}_{jg} = \exp(\beta' x_{jg})$. If the random case-specific effects, $\exp(\eta_{jg})$, are i.i.d. gamma-distributed with parameters $(\delta^{-1}\tilde{\lambda}_{jg}, \delta^{-1}\tilde{\lambda}_{jg})$, with $\delta^{-1} > 0$, then these have an expected value of one and a variance of $\delta^{-1}\tilde{\lambda}_{jg}$. Under these assumptions, the expected probability that an audience member in episode g gives their endowment to candidate j is given by:

$$E(\tilde{p}_{jg}) = \frac{\tilde{\lambda}_{jg}}{\sum_{j=1}^{J} \tilde{\lambda}_{jg}} \tag{3}$$

These expected choice probabilities are not affected by η_{jg}, meaning that the Dirichlet-multinomial model generates the same predictions as the grouped conditional logit model that results if this random component were omitted from Equation 1. However, the grouped conditional logit model fails to account for the correlation between choices arising from unobserved characteristics. Consequently, the donation counts will exhibit greater variability across episodes than the grouped conditional logit model predicts. This misspecification, known as overdispersion, results in distorted standard errors and can thus lead to incorrect conclusions regarding statistical significance. The Dirichlet-multinomial model efficiently accounts for this extra-multinomial variation (Shonkwiler and Hanley, 2003; Guimarães and Lindrooth, 2007).

We obtain the estimates for β by maximizing the log of the following likelihood function, which takes the form of a Dirichlet-multinomial multivariate distribution:

$$L = \prod_{g=1}^{G} \frac{n_g! \, \Gamma(\delta^{-1}\tilde{\lambda}_g)}{\Gamma(\delta^{-1}\tilde{\lambda}_g + n_g)} \prod_{j=1}^{J} \frac{\Gamma(\delta^{-1}\tilde{\lambda}_{jg} + n_{jg})}{\Gamma(\delta^{-1}\tilde{\lambda}_{jg}) n_{jg}!} \tag{4}$$

where n_{jg} denotes the number of audience members in episode g who donate money to candidate j, n_g is the total number of audience members in episode g, and $\tilde{\lambda}_g = \sum_{j=1}^{J} \tilde{\lambda}_{jg}$.

To facilitate interpretation and comparison of the coefficients, we cross-sectionally standardize all continuous independent variables to have a mean of zero and a standard deviation of one.

3.3 Regression results

Table 3 presents the Dirichlet-multinomial regression results. Models 1 through 3 investigate the effect of the impact variables, controlling for candidates' age, gender, and physical attractiveness.[18] When we include the impact variable measuring the well-being effect of €100, a one-standard-deviation increase in this variable raises a candidate's share of the prize pool by 4.0 percentage points (Model 1; $p < 0.001$). When we instead use the impact variable measuring the effect of €10,000, we observe a more sizable effect: a one-standard-deviation increase in this variable raises the share by 7.1 percentage points (Model 2; $p < 0.001$). When both impact variables are included, the effect of the impact of €10,000 remains economically substantial and statistically highly significant (Model 3; $p < 0.001$), whereas the effect of the impact of €100 drops markedly to 1.8 percentage points and becomes only marginally significant ($p = 0.093$).

Model 4 investigates the effects of the variables capturing the four causal attribution dimensions. As expected, candidates who had more control over the cause of their situation received less: a one-standard-deviation increase in personal control reduces the share by 4.6 percentage points ($p = 0.001$). The other CDSII dimensions do not significantly affect the division of the money (all $p > 0.198$).

Models 5 through 7 investigate the effects when both the impact and causal attribution variables are included. The coefficients of the two impact variables and personal control are somewhat smaller than before, but the general picture remains unchanged. The well-being impact of €10,000 has a stronger effect than that of €100 and largely drives out the latter when they are combined in one model, suggesting that the €10,000 version is the better measure. The only dimension of causal attribution that matters is personal control, but in absolute terms its effect size is substantially smaller than that of the impact of €10,000.

Age, gender, and physical attractiveness are insignificant in all models. The same holds for all two-way interactions among these variables, as well as for their three-way interaction (untabulated).

[18] We do not control for candidates' race, as only 10 (6%) are non-white. On average, non-white candidates received a larger share than whites (45.4% vs. 32.6%).

Table 3: Dirichlet-multinomial regression results

The table presents average marginal effects from Dirichlet-multinomial regression analyses. The dependent variable is the proportion of the prize pool that a candidate received. 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. Other definitions are as in Table 2. Standard errors are in parentheses. Asterisks denote statistical significance at the one ***, five **, and ten * percent level, respectively.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Impact €100	0.040***		0.018*		0.033***		0.017*
	(0.011)		(0.010)		(0.010)		(0.010)
Impact €10,000		0.071***	0.064***			0.061***	0.054***
		(0.011)	(0.012)			(0.011)	(0.012)
Personal control				-0.046***	-0.039***	-0.027**	-0.027**
				(0.014)	(0.014)	(0.013)	(0.013)
External control				-0.014	-0.014	-0.018	-0.018
				(0.014)	(0.013)	(0.012)	(0.012)
Internal locus				-0.017	-0.021*	-0.009	-0.012
				(0.013)	(0.013)	(0.012)	(0.012)
Stability				0.003	0.001	0.001	0.000
				(0.015)	(0.014)	(0.014)	(0.014)
Age	0.013	-0.009	-0.006	0.002	0.004	-0.011	-0.009
	(0.013)	(0.012)	(0.012)	(0.013)	(0.012)	(0.012)	(0.012)
Female	0.033	0.005	0.000	0.017	-0.002	-0.012	-0.018
	(0.025)	(0.023)	(0.023)	(0.025)	(0.025)	(0.023)	(0.023)
Attractiveness	0.019	0.002	0.008	0.000	0.010	-0.002	0.004
	(0.013)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)	(0.012)
Log-likelihood	-444.7	-433.7	-432.3	-440.3	-435.6	-428.4	-427.1
Pseudo R^2	0.117	0.286	0.299	0.223	0.281	0.361	0.375
No. candidates	165	165	165	165	165	165	165
No. episodes	55	55	55	55	55	55	55

3.4 Dominance analysis

To evaluate the relative importance of impact and responsibility, we use general dominance analysis. This method decomposes the overall fit of a regression model into the additively separate contributions of each independent variable by computing the average contribution of each variable across all possible combinations of the other variables (Budescu, 1993; Azen and Budescu, 2003). General dominance analysis can be applied to any measure of fit; for our Dirichlet-multinomial regression, we assess the contributions to the log-likelihood, which serves as the objective function.

We conduct the general dominance analysis for Model 6, which includes the impact of €10,000,

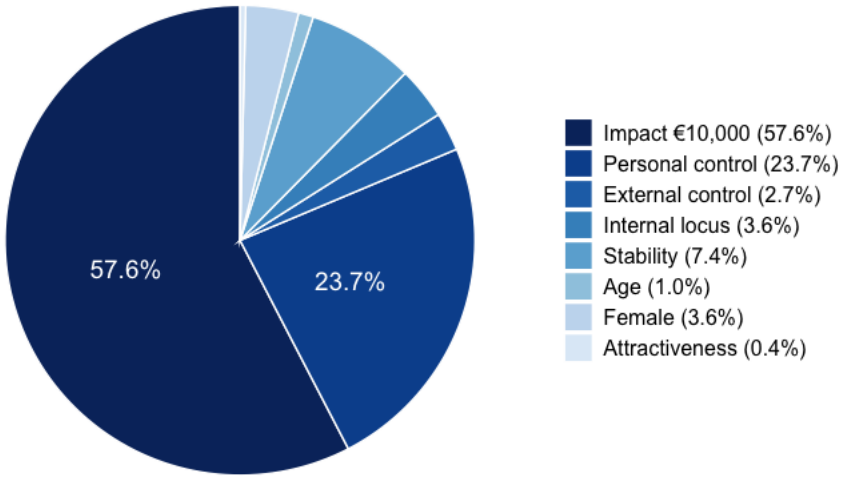

Figure 2: Contributions to model fit. The figure shows the outcomes of the general dominance analysis for Model 6 in Table 3. Slices represent the percentage contribution of the given factor to the improvement of the empirical fit of the model relative to the null model.

personal control, the remaining three causal attribution dimensions, and age, gender, and attractiveness.[19] Figure 2 displays the contributions in a pie chart. Together, impact and personal control account for 81 percent of the model's improvement in fit relative to the null model, with impact contributing 57.6 percent and personal control 23.7 percent. This indicates that the contribution of impact is 2.4 times that of personal control. The remaining variables all add little to the explanatory power.

Based on the analyses up to here, we conclude that both impact and responsibility considerations are important determinants, that giving increases more with impact than it decreases with responsibility, and that the explanatory power of impact is substantially larger than that of responsibility.

4. Robustness Analyses

In this section, we examine the sensitivity of our results to the inclusion of additional control variables (4.1), the exclusion of subsets of raters or candidates (4.2), alternative methodological choices (4.3), and alternative proxies for impact and responsibility (4.4). We also summarize the findings of a replication study (4.5).

[19] A natural approach when comparing the importance of two determinants is to use a single proxy for each. For impact we use the €10,000 rather than the €100 version, as the previous analyses indicated that the former is the better measure. Robustness analyses in Subsection 4.4 show that the flexibility of choosing between two proxies for impact but not for responsibility does not drive our conclusions.

4.1 Additional control variables

First, we investigate the robustness of our results to the inclusion of additional control variables. In Table 4, the first alternative regression model adds the amount of money requested by the candidate, the second adds dummy variables for the candidate's position in the presentation order, and the third adds the duration of the video clip that raters watched (in seconds). These variables were not in the main specifications for different reasons.

The requested amount and the position in the order may partially capture our main variables of interest. Candidates who ask for more money are likely facing more significant issues that are harder to solve without audience donations. Their well-being therefore likely benefits more from resolving their financial need in full (by receiving €10,000) and less from receiving the smallest possible amount (€100). The correlation coefficients in Table C1 in Appendix C support these conjectures. Nevertheless, our main results remain virtually unchanged when we include the requested amount in the regression model. This variable is economically and statistically insignificant, and its inclusion does not materially alter either the other coefficients or the ratio of the contributions of impact and personal control to the model's overall fit.

The order of the presentation of the cases is endogenous. Producers are incentivized to maximize viewer engagement, and may therefore place more interesting cases first or last. Indeed, we observe that the order is significantly related to aspects of the cases. For example, the first candidate on average scores higher on the well-being impact of receiving €10,000 and lower on personal control, and is relatively young (untabulated). The main results remain virtually unchanged when we include dummy variables for the first and third positions. Neither the existing coefficients nor the ratio of the contributions of impact and personal control to the model's overall fit are materially affected. The candidate whose case was featured second received the least money, while the one who was third received the most. The difference of 7.3 percentage points between those two is sizable. This pattern could reflect primacy and recency effects (see, for example, Mantonakis et al., 2009), but the non-random assignment of candidates to positions makes it impossible to draw any definitive conclusions.

The duration of a video clip is, at best, a crude proxy for the total time that was spent on a case in the recording studio, as the panel interviews are edited for broadcast. The variable is nonsignificant, and its inclusion in the model leaves the results for all other variables unchanged.

Table 4: Robustness analyses

The table presents average marginal effects from supplementary regression analyses (Panel A), and the outcomes of the corresponding general dominance analyses (Panel B). All results, except for the column labelled *OLS*, derive from Dirichlet-multinomial regression. In the columns *Amount requested*, *Order*, and *Video duration*, the regression model includes one or two additional control variables. *No familiarity*, *No pairs*, and *No outside beneficiary* exclusively use the 65 raters who declared they had never seen the show before, the cases where there was only one candidate on the stage, and the cases where the candidates themselves or their family were the primary recipients, respectively. *Unstandardized*, *PCF*, and *OLS* report the results without the standardization of (underlying item) scores at the rater level, when the four CDSII dimensions are replaced by factors that derive from principal-component analysis, and when ordinary least squares regression (with episode-level clustered standard errors) is used instead of Dirichlet-multinomial regression, respectively. In the last column, *(Pseudo)* R^2 is the regular, unadjusted R-squared for ordinary least squares regression; in all other columns it is the Efron pseudo R-squared. *First case* and *Last case* are dummy variables indicating whether the candidate's case was featured first or last in the episode, respectively. Other definitions are as in Tables 2 and 3.

		Robustness analyses								
	Model 6	Amount requested	Order	Video duration	No familiarity	No pairs	No outside beneficiary	Un-standardized	PCF	OLS
Panel A: Regression results										
Impact €10,000	0.061***	0.061***	0.061***	0.061***	0.053***	0.059***	0.068***	0.067***	0.062***	0.064***
	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)	(0.013)	(0.012)	(0.015)	(0.011)	(0.015)
Personal control	-0.027**	-0.026*	-0.024*	-0.027**	-0.028**	-0.029*	-0.039***	-0.029*	-0.031***	-0.044***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.016)	(0.014)	(0.016)	(0.010)	(0.016)
External control	-0.018	-0.019	-0.016	-0.018	-0.021*	-0.027*	-0.009	-0.015	-0.016	-0.021
	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.015)	(0.011)	(0.014)	(0.010)	(0.016)
Internal locus	-0.009	-0.009	-0.016	-0.009	-0.007	-0.017	-0.014	-0.007	-0.008	-0.020
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.015)	(0.012)	(0.013)	(0.010)	(0.019)
Stability	0.001	0.001	0.001	0.001	0.004	0.001	-0.010	0.008	0.000	-0.002
	(0.014)	(0.014)	(0.013)	(0.014)	(0.013)	(0.016)	(0.014)	(0.016)	(0.010)	(0.017)
Age	-0.011	-0.010	-0.016	-0.011	-0.007	-0.014	-0.003	-0.011	-0.012	-0.006
	(0.012)	(0.012)	(0.013)	(0.012)	(0.013)	(0.014)	(0.012)	(0.012)	(0.012)	(0.015)
Female	-0.012	-0.012	-0.033	-0.012	0.001	-0.017	-0.019	0.001	-0.014	-0.012
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.026)	(0.023)	(0.023)	(0.023)	(0.033)
Attractiveness	-0.002	-0.001	-0.005	-0.002	-0.005	0.002	0.011	0.003	-0.001	-0.002
	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)	(0.013)	(0.011)	(0.013)	(0.011)	(0.016)
Amount requested		0.007								
		(0.010)								
First case			0.040							
			(0.027)							
Last case			0.073***							
			(0.024)							
Video duration				-0.001						
				(0.010)						
Constant										0.342***
										(0.023)
Log-likelihood	-428.4	-428.2	-423.7	-428.4	-432.8	-352.2	-372.9	-432.6	-428.0	
Mean log-likelihood	-2.596	-2.595	-2.568	-2.596	-2.623	-2.463	-2.437	-2.622	-2.594	
(Pseudo) R^2	0.361	0.369	0.422	0.361	0.314	0.324	0.480	0.325	0.370	0.294
No. candidates	165	165	165	165	165	143	153	165	165	165
No. episodes	55	55	55	55	55	53	55	55	55	55

Table 4: Robustness analyses (continued)

		Robustness analyses								
	Model 6	Amount requested	Order	Video duration	No familiarity	No pairs	No outside beneficiary	Un-standardized	PCF	OLS
Panel B: Dominance analysis results										
Impact €10,000	57.6	56.9	47.1	57.4	53.9	53.7	57.0	51.4	59.1	50.4
Personal control	23.7	22.6	17.9	23.4	25.2	23.9	25.5	28.1	28.5	30.2
External control	2.7	2.8	1.9	2.7	4.2	4.4	0.8	1.9	2.9	2.3
Internal locus	3.6	3.4	4.0	3.6	2.1	4.1	5.7	3.2	2.3	7.5
Stability	7.4	7.2	5.5	7.4	9.1	7.1	5.0	9.0	1.7	6.6
Age	1.0	0.9	1.2	1.0	0.6	1.5	0.4	1.2	1.1	0.5
Female	3.6	3.5	3.2	3.5	4.5	4.0	3.9	4.4	4.0	2.3
Attractiveness	0.4	0.4	0.3	0.4	0.4	1.3	1.6	0.8	0.4	0.2
Amount requested		2.2								
Order			18.9							
Video duration				0.6						
Impact €10,000 / Personal control	2.43	2.51	2.63	2.45	2.14	2.25	2.24	1.83	2.08	1.67

4.2 Sample restrictions

Next, we examine the robustness of our findings to excluding a subset of raters and subsets of candidates. Although most raters reported that they had never seen GMG before, a minority indicated that they had seen at least one episode or were unsure. As shown in Table 4, the results remain practically unchanged when we only consider the responses of the 65 raters who were certain they had never seen the show before. For this subgroup, the coefficient of the impact measure is somewhat smaller than that for all raters, and the ratio of the contributions of impact and personal control is somewhat lower, but the contribution of impact remains more than twice as large as that of personal control.

Most cases featured a single candidate on the stage, and most of the times the candidates themselves or their families were the primary beneficiaries of the money given by the audience. Occasionally, however, a case was represented by two individuals, or a candidate requested money for an outside beneficiary, such as a sports, cultural, or societal association. Our findings are robust to excluding the 20 cases represented by two people, and to excluding the 12 cases where the candidate or their family was not the primary beneficiary. Excluding pairs leaves the coefficients and conclusions virtually unchanged.[20] When cases with outside beneficiaries are excluded, the coefficients for both impact and

[20] Excluding those 20 cases reduces the sample by two additional ones, as the model requires at least two cases per episode, a condition not met for two episodes.

personal control become somewhat more pronounced, but the ratio of the contributions of impact to personal control remains well above two.

4.3 Alternative methodological approaches

Our analysis relied on several methodological choices, including the standardization of item scores at the rater level, the simple averaging of the three items for each CDSII dimension, and the use of Dirichlet-multinomial regression. Here, we discuss three alternative approaches: omitting the standardization, using principal-component analysis to construct the CDSII-based factors, and using standard ordinary least squares (OLS) regression. The last three columns of Table 4 show the results.

Without the standardization of item scores at the rater level, the pseudo R-squared is somewhat lower. The coefficients are largely unchanged, but the relative contribution of impact to the overall empirical fit is lower, suggesting relatively high heterogeneity in how raters interpreted the survey question that uses a percentile scale for measuring the well-being impact of €10,000. Despite this decrease, the contribution of impact to the overall model fit remains much larger than that of personal control.

As discussed in Subsection 2.7, principal-component analysis of the standardized scores for the 12 CDSII items identifies four principal factors that closely correspond to the four CDSII dimensions. As a robustness check, we consider the results that we obtain if we construct each of the four CDSII dimensions in our regression model on the basis of the factor loadings. Consistent with the idea that this data-driven approach leads to better measurement, the estimated effect of personal control becomes somewhat more pronounced. At the same time, impact remains more than twice as important as personal control in explaining the giving behaviour.

The specific structure of our data required the use of Dirichlet-multinomial regression. For completeness, we also present OLS results, with standard errors clustered at the episode level. Although OLS regression does not account for important features of our data, the results are qualitatively similar. The most notable difference is that the coefficient and empirical-fit contribution of personal control are somewhat more pronounced. However, the role of impact remains substantially more important than that of personal control. As expected, OLS explains the data less well: the R-squared is 0.294, compared with the Efron pseudo R-squared of 0.361 for the Dirichlet-multinomial model.[21]

[21] The Efron pseudo R-squared measures the proportion of variance explained by the Dirichlet-multinomial regression, in a similar manner as the regular, unadjusted R-squared in OLS regression. The lower value for the OLS model is noteworthy, as OLS maximizes the R-squared, whereas the Dirichlet-multinomial model maximizes the log-likelihood, not the Efron pseudo R-squared.

Table 5: Robustness analyses with alternative impact and responsibility proxies

The table presents average marginal effects from supplementary Dirichlet-multinomial regression analyses (Panel A), and the outcomes of the corresponding general dominance analyses (Panel B). In the columns labelled *Manageable*, *Regulate*, and *Power*, the variable *Personal control* is exclusively measured on the basis on raters' evaluations of the extent to which the cause of the candidate's situation was manageable by the candidate, the extent to which the candidate could regulate it, or the extent to which the candidate had power over it, respectively. In *Impact measured as average*, the variable *Impact* is calculated as the average of *Impact €100* and *Impact €10,000*; in the other columns this variable is identical to *Impact €10,000*. Other definitions are as in Tables 2 and 3.

		Personal control measured by one item			Impact measured
	Model 6	Manageable	Regulate	Power	as average
Panel A: Regression results					
Impact	0.061***	0.062***	0.063***	0.062***	0.055***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.010)
Personal control	-0.027**	-0.026**	-0.020*	-0.021	-0.030**
	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)
External control	-0.018	-0.018	-0.019	-0.015	-0.017
	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)
Internal locus	-0.009	-0.010	-0.013	-0.007	-0.017
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Stability	0.001	0.003	0.006	0.006	0.000
	(0.014)	(0.013)	(0.013)	(0.013)	(0.014)
Age	-0.011	-0.012	-0.011	-0.012	-0.003
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Female	-0.012	-0.013	-0.009	-0.008	-0.018
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Attractiveness	-0.002	-0.002	-0.002	-0.001	0.010
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Log-likelihood	-428.4	-428.4	-429.1	-429.1	-428.5
Pseudo R^2	0.361	0.362	0.347	0.351	0.359
No. candidates	165	165	165	165	165
No. episodes	55	55	55	55	55
Panel B: Dominance analysis results					
Impact	57.6	58.8	61.9	60.2	54.4
Personal control	23.7	22.2	17.2	20.9	25.1
External control	2.7	2.7	2.9	2.2	2.0
Internal locus	3.6	3.7	4.5	3.5	5.3
Stability	7.4	7.5	8.1	7.9	7.6
Age	1.0	1.1	1.0	1.1	0.3
Female	3.6	3.7	3.9	3.8	3.8
Attractiveness	0.4	0.4	0.4	0.4	1.5
Impact / Personal control	2.43	2.65	3.59	2.88	2.16

4.4 Alternative impact and responsibility proxies

Our comparison of the importance of the roles of impact and responsibility required proxy variables. For responsibility, we used the personal control dimension of the CDSII. For impact, we selected between two self-designed proxies on the basis of their empirical goodness-of-fit. Here, we argue that the flexibility in choosing between two proxies for impact but not for responsibility does not pose a major concern.

First, the personal control measure has its own advantage, because it is an average of three items. Raters assessed the extent to which the cause of the candidate's situation was manageable by the candidate (item 2; see Table 1), the extent to which the candidate could regulate it (item 4), and the extent to which the candidate had power over it (item 10). As each of these items is designed to gauge personal control, averaging them is likely to reduce measurement error. Reduced measurement error lessens the attenuation of a variable's coefficient and increases its contribution to the model fit. The impact measure lacks this benefit, because it derives from a single question.

Second, we obtain similar results when we treat the three items as separate measures of personal control and select the one that performs best, analogous to our approach for impact. Table 5 presents the three sets of results when we measure personal control by the individual items. The manageability variant yields the best fit. Using this item as our measure of personal control produces results that are similar to those obtained previously. If anything, the contribution of impact relative to that of personal control to the model fit is larger than before.

Last, we also obtain similar results when we construct the impact measure as an average of multiple items, analogous to the construction of the personal control measure. Using the average of the impact of €100 and €10,000—instead of the impact of €10,000—slightly reduces the coefficient of impact and its relative contribution to the model fit, and slightly amplifies those values for personal control. Nevertheless, impact remains more than twice as important as personal control in explaining the giving behaviour of the audience.

4.5 Replication study

We conclude the series of robustness analyses with a summary of the findings of a replication study. A comprehensive account of the experimental design and results is provided in Lin et al. (2026), a companion to the present paper.[22]

[22] The experiment for the replication study was conducted as a behavioral pilot study for a neuroimaging project that will use GMG to investigate the neural underpinnings of giving behavior. Its purpose was twofold: to

Lin et al. consider the subset of 28 episodes in which every case was represented on the stage by one candidate only (no pairs) and in which candidates themselves or their family were always the primary recipients (no outside beneficiaries). In rating sessions, subjects evaluated five pseudo-randomly selected cases each, without regard for episode structure. A total of 441 Dutch-speaking student-subjects provided an average of 26 ratings per item for each of the 84 cases.

For impact, the replication employed a simplified measure: raters were asked to assess the extent to which receiving the requested amount would improve the candidate's well-being, on a 10-point Likert scale.[23] Personal control was again assessed using the three personal-control dimension items of the CDSII. In addition, Lin et al. measured raters' willingness to donate by asking them how likely they would be to donate €100 to the candidate if they were a member of the studio audience, on a continuous scale running from 0 ("not at all likely") to 100 percent ("very likely").

Owing to the much larger number of raters per case, the ICCs for impact and personal control in the replication study are considerably higher than those for the original data (for the 28 episodes; impact: 0.89 vs. 0.40, personal control: 0.94 vs. 0.69). The correlation between the original measure for the impact of €10,000 and the new, simplified one is 0.55, which is relatively high in light of the differences between the two measures and in light of the low ICC of the original measure. The original and new personal control scores are also strongly correlated, with a coefficient of 0.74.

Table 6 summarizes the results of the replication study. For comparison, the first model shows the main results for the subset of 28 episodes, based on the original data.[24] The second model uses the new impact and personal control measures, while controlling for the same set of covariates as in the original analysis. With the new measures, the effect of impact is somewhat larger, while that of personal control decreases markedly and becomes statistically nonsignificant. Dominance analysis, however, shows that when considering the average contributions of the two factors across all possible submodels, the ratio of their contributions to the empirical fit is barely affected.

A potential concern is that student-raters may hold different views than the audience members in the studio. The replication study addresses this concern by examining not only the donation behaviour of the audience but also the donation propensities of the individual raters themselves. Reassuringly,

replicate the key findings of the present paper, and to gather supplementary data to inform stimulus selection and experimental design. Two of the authors of the present paper also contributed to the replication study.

[23] This alternative measure was used to reduce raters' cognitive load. Theoretically, however, its limited focus on the well-being effect for the candidate only, as well as the use of the case-specific requested amount rather than a fixed amount, align less well with the aim of capturing the normative utilitarian perspective.

[24] Compared with the results for all episodes (see Table 3, Model 6), the role of personal control is more pronounced, but impact remains a substantially more important contributor to the explanatory power than personal control.

Table 6: Main results replication study

The table summarizes the replication results reported in Lin et al. (2026), for the 28 episodes in which each case was represented by one candidate on the stage and where candidates themselves or their family were always the primary recipients. Definitions are as in Tables 2 and 3 in Lin et al. In the first two models, the dependent variable is the candidate's share of the prize pool, and the reported results are average marginal effects from Dirichlet-multinomial regression analyses. In the last two models, the dependent variable is the stated donation propensity of a rater towards a particular candidate, and the reported results are OLS coefficients. The control variables are *External control*, *Internal locus*, *Stability*, *Age*, *Female*, and *Attractiveness*. Standard errors are in parentheses. Asterisks denote statistical significance at the one ***, five **, and ten * percent level, respectively.

	Audience donations		Individual rater propensities	
	Original data	New impact and personal control data	Without fixed effects	With fixed effects
Panel A: Regression results				
Impact	0.081***	0.095***	0.468***	0.386***
	(0.015)	(0.019)	(0.023)	(0.023)
Personal control	-0.055***	-0.010	-0.289***	-0.240***
	(0.018)	(0.025)	(0.027)	(0.024)
Control variables	Y	Y	Y	N
Candidate FEs	N	N	N	Y
Log-likelihood	-212.9	-205.5	-	-
(Pseudo) R^2	0.544	0.642	0.459	0.548
No. ratings	-	-	2,170	2,170
No. raters	-	-	434	434
No. candidates	84	84	84	84
No. episodes	28	28	-	-
Panel B: Dominance analysis results				
Impact	48.8	53.1	56.2	57.0
Personal control	27.1	27.3	30.1	35.5
Controls	24.1	19.6	13.7	7.5
Impact / Personal control	1.80	1.94	1.87	1.61

these stated propensities of raters closely mirror the behaviour of the audience members on the show: regressing the candidate's share of the prize pool on raters' average standardized donation propensity—using the Dirichlet–multinomial model—yields a comparatively high Efron pseudo-R-squared of 0.652.

The remaining two models in Table 6 consider the relationship between raters' individual assessments of impact and personal control and their own stated donation propensities. The third model uses the standard set of control variables, the fourth instead uses fixed effects to eliminate systematic differences in raters' donation propensities across candidates.[25] Both sets of results confirm that impact and personal control are strong predictors, and that impact is the more important of the two.

[25] The fourth model thus exclusively shows how between-rater variation in evaluations of a given case explain between-rater differences in their propensity to donate to that case.

Additionally, Lin et al. find similar results for male and female raters, providing further support for the generality of the observed donation patterns.

Taken together, these replication results reinforce the robustness of our findings: both impact and personal control shape giving behaviour, with impact being the more influential factor. Using new ratings and an alternative impact measure, the study replicates this pattern not only for the donations by the audience, but also for raters' own (stated) willingness to donate.

5. Conclusion and Discussion

This paper examines the relative importance of perceptions of impact and responsibility for giving decisions. From a utilitarian perspective, people should prioritize the well-being benefits of their donations ("impact"). In contrast, the fairness principles of equity and desert imply that people should be more inclined to give if the recipient's need arises from misfortune rather than from factors within their control ("responsibility").

Our data are from a TV show in which an audience of one hundred people divides ten thousand euros among three candidates who face a problem beyond their financial capacity. Independent raters evaluated various attributes of the candidates and their predicaments, including the well-being impact of donations and the degree of control the candidates had over the cause of their situation.

We find that donations increase strongly with impact and decrease moderately with responsibility, and that the contribution of impact to the fit of our regression models is approximately twice that of responsibility. We find no evidence of discrimination on the basis of age, gender, or physical attractiveness. These conclusions are robust across a wide range of analytical choices and are confirmed by a replication study.

The greater importance of utilitarian considerations over equity and desert considerations in our setting contrasts with the picture emerging from the literature. Empirical research generally shows strong evidence for equity and desert, while the evidence for utilitarianism is mixed. Charity donations, for example, are often only weakly influenced by the impact they have (Hsee and Rottenstreich, 2004; Null, 2011; Hasford et al., 2015; Karlan and Wood, 2017; Berman et al., 2018).

A possible explanation for this contrast is that in our TV show, audience members get a detailed glimpse into the lives and problems of the candidates. As a result, they can easily imagine the impact their donation would have on the candidates' lives and well-being. Such vivid information is typically absent in laboratory experiments and in real-world charitable giving. Indeed, several studies show that the importance of impact as a driver of donation decisions increases when people receive clearer effectiveness information (Caviola et al., 2014; Metzger and Günther, 2019; Caviola et al., 2020;

Kretschmer et al., 2025). In this light, our results suggest that providing concrete evidence of the tangible, positive outcomes of donations can shift donor focus toward utilitarian considerations. By clearly showing how donations affect beneficiaries, charities and crowdfunders can steer donors toward decisions that maximize well-being, and enhance the overall effectiveness of giving.

To the best of our knowledge, the present study is the first to directly compare the descriptive validity of utilitarianism with that of equity- and desert-based principles in giving behaviour. Our finding that judgments of impact and personal control are correlated underscores the importance of examining the two perspectives simultaneously. Individuals perceived as having less control over their situation are assigned higher impact scores, and failing to account for one can thus lead to an overestimation of the effect of the other.

For utilitarianism, we measured the perceived impact of both €100—the amount each audience member individually donates—and €10,000—the maximum a candidate could receive and the amount that typically fully covers their financial need. Our findings show that the perceived impact of €10,000 is a stronger predictor than that of €100. A possible explanation for the limited explanatory power of the €100 measure is that such a relatively small gift does not materially help any of the candidates in resolving their problems.[26] Raters' low average assessment of the well-being benefit of €100 corroborates this explanation. The strong explanatory power of the €10,000 measure suggests that audience members asked themselves which candidate would benefit most if their financial need were addressed, anticipating that others in the audience would reason similarly.

In light of findings from past research, the absence of evidence for discrimination on physical attractiveness in our setting may seem surprising. For example, attractive people tend to receive higher wages (Hamermesh and Biddle, 1994; Mobius and Rosenblat, 2006; Graham et al., 2017; Halford and Hsu, 2020), and are more likely to be elected (Hamermesh, 2006; Rosar et al., 2008, Berggren et al., 2010). In the domain of charitable giving, Landry et al. (2006) find that gift sizes in a door-to-door fundraising campaign increase with the physical attractiveness of female solicitors. Similarly, early research suggests that women are more likely to receive help if they are seen as more attractive (Mims et al., 1975; West and Brown, 1975). At the same time, however, there is also work that finds evidence of a beauty penalty rather than a premium (Fisher and Ma, 2014; Yu et al., 2024). One possible reason why there is no attractiveness effect in the present study is that the audience members are provided with and implicitly asked to compare an abundance of information about the candidates. Indeed, Cryder et al. (2017) find that deliberation steers giving choices away from

[26] The explanatory power of the impact of €100 may have been larger if we had measured it as its contribution to well-being in addition to that of the expected donations of others, rather than in isolation. However, incorporating beliefs about others' donations into the measure would have made the survey question too complicated.

attractive recipients. Consistent with this, Pope and Sydnor (2011) and Duarte et al. (2012) find little role for borrowers' perceived attractiveness in an online peer-to-peer lending market characterized by substantial information availability for lenders.

While donations are often influenced by a broader range of factors than those examined here (for a literature review, see, for example, Bekkers and Wiepking, 2011, and Andreoni and Payne, 2013), many are invariant across the giving decisions that are made in the controlled setting of our TV show. This allows a relatively clean comparison of the importance of impact and responsibility considerations, with limited scope for alternative explanations and minimal noise deriving from unaccounted factors. For instance, since all candidates are equally identifiable and since there are no overhead costs, neither the identifiable victim effect (Jenni and Loewenstein, 1997; Small and Loewenstein, 2003; Small et al., 2007) nor overhead aversion (Baron and Szymanska, 2011; Caviola et al., 2014; Gneezy et al., 2014) can account for how the audience members choose between the three candidates. Likewise, decision makers cannot keep the money for themselves, and differences in their general tendencies to give can therefore neither be of influence. Last, donors are endowed with money provided by a third party, to which none of the candidates has made any contribution. This allows us to examine the effect of candidates' responsibility for their situation of need in isolation, without the confounding influence of differences in their responsibility for generating the resources. If candidates had earned the pool of money themselves, the donation decisions of the audience members would likely be influenced by perceptions of entitlements arising from such contributions, as well as by attitudes towards the protection of entitlements (Bauer et al., 2022).

Another benefit of our setting is that it closely aligns with the ideal of the impartial spectator design that is often used in laboratory studies on moral preferences (Konow, 2009): the decision makers in the audience are anonymous and endowed with money that they must donate to one of three candidates with whom they have no personal connection. Also, our decision makers are confronted with a rich variety of candidates facing real-world financial challenges, and the stakes are substantial—both in terms of consequences for the candidates and relative to those typically used in laboratory experiments.

A difference between our setting and the standard impartial spectator design is that our audience members do not act as sole decision makers who individually determine the entire allocation. Instead, each has only a small marginal influence on the overall distribution, which is co-determined by the choices of 99 others. Theoretically, this setup introduces strategic considerations: an audience member's donation choice should not only depend on their preferences regarding the overall allocation, but also on their beliefs about the choices of others. For instance, if an audience member expects that a sufficiently large majority will support the candidate they believe should receive most

of the funds, they may decide to allocate their share to their second-ranked candidate instead. The extent to which such strategic considerations have had an influence is unclear. Nevertheless, in equilibrium, the observed distributions should still reflect the moral preferences of the broader population.

The strategic aspect of our setting is common in many real-world giving situations, particularly in charitable giving, where donors typically contribute alongside many others. Future research could investigate how people think about their own contributions and those of others, and how they assess the impact of giving in such contexts. Our finding that the perceived well-being impact of €10,000 is a strong determinant suggests that individuals tend to take a collective perspective.

Using a TV show as the setting for empirical research into real-world decision making offers unique advantages, but can also raise possible concerns (Larkin and Pines, 2003; Reback and Stowe, 2011; Baltussen et al., 2016). One relates to external validity, because decisions in TV shows are typically made under the pressure of the limelight. In our show, however, the decision makers are part of the audience and rarely visible on camera. Their choices are as good as anonymous, as only the overall distribution of donations is made public, not their individual decisions. As a result, they do not face the stresses and pressures that TV show participants commonly face. Another concern are possible selection effects. With many TV shows, people will be more likely to self-select into auditions if they are comfortable with the idea of being in the limelight, and producers often select participants on the basis of their personality and skills. In GMG, these selection influences are largely absent. On the surface, our audience resembles a broad (middle-class) cross-section of the general population. Of course, audience members must have had the time and interest to travel to the studio and spend half a day there, but it seems unlikely that this has materially influenced our results in a specific direction. Indeed, the replication study finds similar results for the stated donation propensities of student-subjects in the laboratory, suggesting that our results are not driven by limelight pressures or audience-selection effects.

A potential concern with our approach is that the measures of impact and personal control reflect the judgments of student-raters, while the giving decisions that we analyse were made by audience members in a television studio. This disconnect could be problematic if the two groups hold markedly different views on how candidates rank in terms of the well-being impact of a donation and their responsibility for their predicaments. At the same time, there is also an upside to the disconnect: because the ratings and the choice data stem from two different sources, we can rule out that the correlation between a particular measure and giving behaviour is driven by common-method bias (Campbell and Fiske, 1959; Podsakoff et al., 2003; Podsakoff et al., 2012). Without this disconnect,

audience members might have tried to act in accordance with their ratings or to provide ratings that are consistent with their choices.

The strong correlations between our measures and the giving decisions suggest that the disconnect between student raters and studio audience is not a major issue. Moreover, the replication study provides further reassurance. First, it reproduced our results using a different group of raters, who independently assessed impact and responsibility. Second, it additionally measured raters' own willingness to donate and found similar results for these stated donation propensities, that is, in the absence of any disconnect. Last, it found a high correlation between raters' donation propensities and the donations of the audience.

We conclude with some suggestions for further research. A limitation of our study is that the aggregate nature of our data prevents us from exploring heterogeneity in people's adherence to normative views on giving. Overall, in our setting utilitarian considerations are more important than equity and desert considerations, but it may well be that many individuals place greater weight on the latter than on the former. Future research could investigate such individual differences, as well as their causes and correlates. The replication study finds no evidence of gender differences in this respect. Another potential factor to consider is cultural background, as moral views have been found to differ across countries (Schokkaert and Devooght, 2003; Gaertner and Schwettmann, 2007; Awad et al., 2018; Almås et al., 2020; Atari et al., 2020; Cappelen et al., 2025).

The relative importance of impact and responsibility considerations may also depend on study-specific features. For example, our results might have been different if people were volunteering their time rather than donating money, if they gave their own money rather than that of a third party, if they were identifiable and accountable for their choices rather than as good as anonymous, or if the donated amounts were less consequential for the recipients. Further research may replicate our analysis under different conditions, or directly investigate the moderating role of such factors. Similarly, new studies may employ alternative operationalizations of utilitarian versus equity and desert considerations.

Lastly, future research could examine how people form their subjective impact and responsibility judgments, and how well these align with more objective benchmarks. For impact, for example, we elicited holistic assessments of the well-being impact of donations, and it would be interesting to explore how people weigh different aspects of a case, such as the number of beneficiaries, the domain and time horizon of the benefits, and the extent to which those benefits are psychological or physical.

References

Adena, Maja, Jeyhun Alizade, Frauke Bohner, Julian Harke, and Fabio Mesters. 2019. "Quality certification for nonprofits, charitable giving, and donor's trust: Experimental evidence." *Journal of Economic Behavior & Organization* 159: 75-100.

Aknin, Lara B., Elizabeth W. Dunn, Ashley V. Whillans, Adam M. Grant, and Michael I. Norton. 2013. "Making a difference matters: Impact unlocks the emotional benefits of prosocial spending." *Journal of Economic Behavior & Organization* 88: 90-95.

Alesina, Alberto, and Eliana La Ferrara. 2005. "Preferences for redistribution in the land of opportunities." *Journal of Public Economics* 89(5-6): 897-931.

Alesina, Alberto, and Paola Giuliano. 2011. "Preferences for redistribution." In *Handbook of Social Economics*, ed. Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, 93-131. North-Holland.

Almås, Ingvild, Alexander W. Cappelen, and Bertil Tungodden. 2020. "Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians?" *Journal of Political Economy* 128(5): 1753-1788.

Andreoni, James, and A. Abigail Payne. 2013. "Charitable giving." In *Handbook of Public Economics* 5, ed. Alan J. Auerbach, Raj Chetty, Martin Feldstein, and Emmanuel Saez, 1-50. Elsevier.

Antonovics, Kate, Peter Arcidiacono, and Randall Walsh. 2005. "Games and discrimination: Lessons from The Weakest Link." *Journal of Human Resources* 40(4): 918-947.

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.

Awad, Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. 2018. "The moral machine experiment." *Nature* 563: 59-64.

Azen, Razia, and David V. Budescu. 2003. "The dominance analysis approach for comparing predictors in multiple regression." *Psychological Methods* 8(2): 129-148.

Baltussen, Guido, Martijn J. van den Assem, and Dennie van Dolder. 2016. "Risky choice in the limelight." *Review of Economics and Statistics* 98(2): 318-332.

Baron, Jonathan, and Ewa Szymanska. 2011. "Heuristics and biases in charity." In *The Science of Giving: Experimental Approaches to the Study of Charity*, ed. Daniel M. Oppenheimer and Christopher Y. Olivola, 215-235. Psychology Press.

Bauer, Alexander Max, Frauke Meyer, Jan Romann, Mark Siebel, and Stefan Traub. 2022. "Needs, equity, and accountability; Evidence on third-party distribution decisions from a vignette study." *Social Choice and Welfare* 59: 769-814.

Bekkers, René, and Pamala Wiepking. 2011. "A literature review of empirical studies of philanthropy: Eight mechanisms that drive charitable giving." *Nonprofit and Voluntary Sector Quarterly* 40(5): 924-973.

Belot, Michèle, V. Bhaskar, and Jeroen van de Ven. 2010. "Promises and cooperation: Evidence from a TV game show." *Journal of Economic Behavior & Organization* 73(3): 396-405.

Belot, Michèle, V. Bhaskar, and Jeroen van de Ven. 2012. "Can observers predict trustworthiness?" *Review of Economics and Statistics* 94(1): 246-259.

Bennett, Randall W., and Kent A. Hickman. 1993. "Rationality and The Price Is Right." *Journal of Economic Behavior & Organization* 21(1): 99-105.

Berggren, Niclas, Henrik Jordahl, and Panu Poutvaara. 2010. "The looks of a winner: Beauty and electoral success." *Journal of Public Economics* 94(1-2): 8-15.

Berk, Jonathan B., Eric Hughson, and Kirk Vandezande. 1996. "The price is right, but are the bids? An investigation of rational decision theory." *American Economic Review* 86(4): 954-970.

Berman, Jonathan Z., Alixandra Barasch, Emma E. Levine, and Deborah A. Small. 2018. "Impediments to effective altruism: The role of subjective preferences in charitable giving." *Psychological Science* 29(5): 834-844.

Betancourt, Hector. 1990. "An attribution-empathy model of helping behavior: Behavioral intentions and judgments of help-giving." *Personality and Social Psychology Bulletin* 16(3): 573-591.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement error in survey data." In *Handbook of Econometrics* 5, ed. James J. Heckman and Edward Leamer, 3705-3843. Elsevier.

Budescu, David V. 1993. "Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression." *Psychological Bulletin* 114(3): 542-551.

Buser, Thomas, Martijn J. van den Assem, and Dennie van Dolder. 2023. "Gender and willingness to compete for high stakes." *Journal of Economic Behavior & Organization* 206: 350-370.

Campbell, Donald T., and Donald W. Fiske. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix." *Psychological Bulletin* 56(2): 81-105.

Cappelen, Alexander W., Ranveig Falch, Zhongjing Huang, and Bertil Tungodden. 2025. "Acceptance of inequality between children: Large-scale experimental evidence from China and Norway." *Economic Journal* 135(667): 999-1020.

Cappelen, Alexander W., Astri Drange Hole, Erik Ø. Sørensen, and Bertil Tungodden. 2007. "The pluralism of fairness ideals: An experimental approach." *American Economic Review* 97(3): 818-827.

Cappelen, Alexander W., James Konow, Erik Ø. Sørensen, and Bertil Tungodden. 2013. "Just luck: An experimental study of risk-taking and fairness." *American Economic Review* 103(4): 1398-1413.

Cappelen, Alexander W., Johanna Mollerstrom, Bjørn-Atle Reme, and Bertil Tungodden. 2022. "A meritocratic origin of egalitarian behaviour." *Economic Journal* 132(646): 2101-2117.

Cappelen, Alexander W., Erik Ø. Sørensen, and Bertil Tungodden. 2010. "Responsibility for what? Fairness and individual responsibility." *European Economic Review* 54(3): 429-441.

Cappelen, Alexander W., and Bertil Tungodden. 2009. "Rewarding effort." *Economic Theory* 39(3): 425-441.

Caviola, Lucius, Nadira Faulmüller, Jim A. C. Everett, Julian Savulescu, and Guy Kahane. 2014. "The evaluability bias in charitable giving: Saving administration costs or saving lives?" *Judgment and Decision Making* 9(4): 303-315.

Caviola, Lucius, Stefan Schubert, and Jason Nemirow. 2020. "The many obstacles to effective giving." *Judgment and Decision Making* 15(2): 159-172.

Cettolin, Elena, and Arno Riedl. 2017. "Justice under uncertainty." *Management Science* 63(11): 3739-3759.

Cohn, Alain, Lasse J. Jessen, Marko Klašnja, and Paul Smeets. 2023. "Wealthy Americans and redistribution: The role of fairness preferences." *Journal of Public Economics* 225, 104977.

Cryder, Cynthia, Simona Botti, and Yvetta Simonyan. 2017. "The charity beauty premium: Satisfying donors' "want" versus "should" desires." *Journal of Marketing Research* 54(4): 605-618.

Dickinson, David L., and Jill Tiefenthaler. 2002. "What is fair? Experimental evidence." *Southern Economic Journal* 69(2): 414-428.

Duarte, Jefferson, Stephan Siegel, and Lance Young. 2012. "Trust and credit: The role of appearance in peer-to-peer lending." *Review of Financial Studies* 25(8): 2455-2484.

Fisher, Robert J., and Yu Ma. 2014. "The price of being beautiful: Negative effects of attractiveness on empathy for children in need." *Journal of Consumer Research* 41(2): 436-450.

Fong, Christina M. 2001. "Social preferences, self-interest, and the demand for redistribution." *Journal of Public Economics* 82(2): 225-246.

Fong, Christina M. 2007. "Evidence from an experiment on charity to welfare recipients: Reciprocity, altruism and the empathic responsiveness hypothesis." *Economic Journal* 117(522): 1008-1024.

Fong, Christina M., and Felix Oberholzer-Gee. 2011. "Truth in giving: Experimental evidence on the welfare effects of informed giving to the poor." *Journal of Public Economics* 95(5-6): 436-444.

Gaertner, Wulf, and Lars Schwettmann. 2007. "Equity, responsibility and the cultural dimension." *Economica* 74(296): 627-649.

Gangadharan, Lata, Philip J. Grossman, Lingbo Huang, C. Matthew Leister, and Erte Xiao. 2023. "Persuadable or dissuadable altruists? The impact of information of recipient characteristics on giving." *Economic Journal* 133(656): 2925-2948.

Gertner, Robert. 1993. "Game shows and economic behavior: Risk-taking on Card Sharks." *Quarterly Journal of Economics* 108(2): 507-521.

Gneezy, Uri, Elizabeth A. Keenan, and Ayelet Gneezy. 2014. "Avoiding overhead aversion in charity." *Science* 346(6209): 632-635.

Gordon, Teresa P., Cathryn L. Knock, and Daniel G. Neely. 2009. "The role of rating agencies in the market for charitable contributions: An empirical test." *Journal of Accounting and Public Policy* 28(6): 469-484.

Guimarães, Paulo, and Richard C. Lindrooth. 2007. "Controlling for overdispersion in grouped conditional logit models: A computationally simple application of Dirichlet-multinomial regression." *Econometrics Journal* 10(2): 439-452.

Graham, John R., Campbell R. Harvey, and Manju Puri. 2017. "A corporate beauty contest." *Management Science* 63(9): 3044-3056.

Greitemeyer, Tobias, and Udo Rudolph. 2003. "Help giving and aggression from an attributional perspective: Why and when we help or retaliate." *Journal of Applied Social Psychology* 33(5): 1069-1087.

Gurevich, Gregory, and Doron Kliger. 2013. "The Manipulation: Socio-economic decision making." *Journal of Economic Psychology* 39: 171-184.

Halford, Joseph T., and Hung-Chia S. Hsu. 2020. "Beauty is wealth: CEO attractiveness and firm value." *Financial Review* 55(4): 529-556.

Hamermesh, Daniel S., and Jeff E. Biddle. 1994. "Beauty and the labor market." *American Economic Review* 84(5): 1174-1194.

Hamermesh, Daniel S. 2006. "Changing looks and changing "discrimination": The beauty of economists." *Economics Letters* 93(3): 405-412.

Hasford, Jonathan, Adam Farmer, and Stacie F. Waites. 2015. "Thinking, feeling, and giving: The effects of scope and valuation on consumer donations." *International Journal of Research in Marketing* 32(4): 435-438.

Hausman, Jerry A. 2001. "Mismeasured variables in econometric analysis: Problems from the right and problems from the left." *Journal of Economic Perspectives* 15(4): 57-67.

Heider, Fritz. 1958. *The Psychology of Interpersonal Relations*. John Wiley & Sons.

Hsee, Christopher K., and Yuval Rottenstreich. 2004. "Music, pandas, and muggers: On the affective psychology of value." *Journal of Experimental Psychology: General* 133(1): 23-30.

Hogarth, Robin M., Natalia Karelaia, and Carlos Andrés Trujillo. 2012. "When should I quit? Gender differences in exiting competitions." *Journal of Economic Behavior & Organization* 83(1): 136-150.

Jenni, Karen, and George Loewenstein. 1997. "Explaining the identifiable victim effect." *Journal of Risk and Uncertainty* 14(3): 235-257.

Jetter, Michael, and Jay K. Walker. 2018. "The gender of opponents: Explaining gender differences in performance and risk-taking?" *European Economic Review* 109: 238-256.

Jones, Edward E., David E. Kanouse, Harold H. Kelley, Richard E. Nisbett, Stuart Valins, and Bernard Weiner. 1972. *Attribution: Perceiving the Causes of Behavior*. General Learning Press.

Kaiser, Henry F. 1958. "The varimax criterion for analytic rotation in factor analysis." *Psychometrika* 23(3): 187-200.

Karlan, Dean, and Daniel H. Wood. 2017. "The effect of effectiveness: Donor response to aid effectiveness in a direct mail fundraising experiment." *Journal of Behavioral and Experimental Economics* 66: 1-8.

Kim, Jae-On, and Charles W. Mueller. 1978. *Factor Analysis: Statistical Methods and Practical Issues*. Sage.

Klein Teeselink, Bouke, Dennie van Dolder, Martijn J. van den Assem, and Jason D. Dana. 2024. "High-stakes failures of backward induction." *Games and Economic Behavior*, forthcoming.

Konow, James. 1996. "A positive theory of fairness." *Journal of Economic Behavior & Organization* 31(1): 13-35.

Konow, James. 2000. "Fair shares: Accountability and cognitive dissonance in allocation decisions." *American Economic Review* 90(4): 1072-1091.

Konow, James. 2001. "Fair and square: The four sides of distributive justice." *Journal of Economic Behavior & Organization* 46(2): 137-164.

Konow, James. 2003. "Which is the fairest one of all? A positive analysis of justice theories." *Journal of Economic Literature* 41(4): 1188-1239.

Konow, James. 2005. "Blind spots: The effects of information and stakes on fairness bias and dispersion." *Social Justice Research* 18(4): 349-390.

Konow, James. 2009. "Is fairness in the eye of the beholder? An impartial spectator analysis of justice." *Social Choice and Welfare* 33(1): 101-127.

Krawczyk, Michał. 2010. "A glimpse through the veil of ignorance: Equality of opportunity and support for redistribution." *Journal of Public Economics* 94(1-2): 131-141.

Kretschmer, Janek, David Reinstein, and Paul Smeets. 2025. "When does charity effectiveness matter to donors? The role of ratings and expectations about cost-effectiveness." Working paper, available at https://ssrn.com/abstract=5207397.

Landry, Craig E., Andreas Lange, John A. List, Michael K. Price, and Nicholas G. Rupp. 2006. "Toward an understanding of the economics of charity: Evidence from a field experiment." *Quarterly Journal of Economics* 121(2): 747-782.

Landry, Craig E., Andreas Lange, John A. List, Michael K. Price, and Nicholas G. Rupp. 2010. "Is a donor in hand better than two in the bush? Evidence from a natural field experiment." *American Economic Review* 100(3): 958-983.

Larkin, Judith E., and Harvey A. Pines. 2003. "Gender and risk in public performance." *Sex Roles* 49(5-6): 197-210.

Levitt, Steven D. 2004. "Testing theories of discrimination: Evidence from Weakest Link." *Journal of Law and Economics* 47(2): 431-453.

Lin, Ting-Yi, Dennie van Dolder, Martijn J. van den Assem, Alexander G. Genevsky, Ale Smidts, and Maarten A. S. Boksem. 2026. "Impact and responsibility in giving behavior: A replication and extension." Working paper, available at https://ssrn.com/abstract=5325790.

List, John A. 2004. "Young, selfish and male: Field evidence of social preferences." *Economic Journal* 114(492): 121-149.

List, John A. 2006. "Friend or Foe? A natural experiment of the prisoner's dilemma." *Review of Economics and Statistics* 88(3): 463-471.

MacAskill, William. 2015. *Doing Good Better: Effective Altruism and a Radical New Way to Make a Difference*. Guardian Faber Publishing.

Mantonakis, Antonia, Pauline Rodero, Isabelle Lesschaeve, and Reid Hastie. 2009. "Order in choice: Effects of serial position on preferences." *Psychological Science* 20(11): 1309-1312.

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.

McFadden, Daniel. 1974. "Conditional logit analysis of qualitative choice behavior." In *Frontiers in Econometrics*, ed. Paul Zarembka, 105-142. Academic Press.

McGraw, Kenneth O., and S.P. Wong. 1996. "Forming inferences about some intraclass correlation coefficients." *Psychological Methods* 1(1): 30-46.

Melkonyan, Tigran, Zvi Safra, and Sinong Ma. 2021. "Justice in an uncertain world: Evidence on donations to cancer research." *Journal of Risk and Uncertainty* 62(3): 281-311.

Metrick, Andrew. 1995. "A natural experiment in Jeopardy!" *American Economic Review* 85(1): 240-253.

Metzger, Laura, and Isabel Günther. 2019. "Making an impact? The relevance of information on aid effectiveness for charitable giving. A laboratory experiment." *Journal of Development Economics* 136: 18-33.

Mims, Patricia R., John J. Hartnett, and W. Robert Nay. 1975. "Interpersonal attraction and help volunteering as a function of physical attractiveness." *Journal of Psychology* 89(1): 125-131.

Mobius, Markus, M., and Tanya S. Rosenblat. 2006. "Why beauty matters." *American Economic Review* 96(1): 222-235.

Mollerstrom, Johanna, Bjørn-Atle Reme, and Erik Ø. Sørensen. 2015. "Luck, choice and responsibility—An experimental study of fairness views." *Journal of Public Economics* 131: 33-40.

Null, Clair. 2011. "Warm glow, information, and inefficient charitable giving." *Journal of Public Economics* 95(5-6): 455-465.

Osborne, Danny, and Bernard Weiner. 2015. "A latent profile analysis of attributions for poverty: Identifying response patterns underlying people's willingness to help the poor." *Personality and Individual Differences* 85: 149-154.

Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff. 2003. "Common method biases in behavioral research: A critical review of the literature and recommended remedies." *Journal of Applied Psychology* 88(5): 879-903.

Podsakoff, Philip M., Scott B. MacKenzie, and Nathan P. Podsakoff. 2012. "Sources of method bias in social science research and recommendations on how to control it." *Annual Review of Psychology* 63: 539-569.

Pope Devin G., and Justin R. Sydnor. 2011. "What's in a picture? Evidence of discrimination from Prosper.com." *Journal of Human Resources* 46(1): 53-92.

Post, Thierry, Martijn J. van den Assem, Guido Baltussen, and Richard H. Thaler. 2008. "Deal or no deal? Decision making under risk in a large-payoff game show." *American Economic Review* 98(1): 38-71.

Reback, Charles, and Kristin Stowe. 2011. "Unnatural experiments: The case of television game shows." *Applied Economics Letters* 18(10): 919-923.

Reisenzein, Rainer. 1986. "A structural equation analysis of Weiner's attribution-affect model of helping behavior." *Journal of Personality and Social Psychology* 50(6): 1123-1133.

Robson, Michael, Owen O'Donnell, and Tom Van Ourti. 2025. "Responsibility-sensitive welfare weights for health." *Journal of Health Economics* 102: 103018.

Roemer, John E., and Alain Trannoy. 2016. "Equality of opportunity: Theory and measurement." *Journal of Economic Literature* 54(4): 1288-1332.

Rosar, Ulrich, Markus Klein, and Tilo Beckers. 2008. "The frog pond beauty contest: Physical attractiveness and electoral success of the constituency candidates at the North Rhine-Westphalia state election of 2005." *European Journal of Political Research* 47(1): 64-79.

Rudolph, Udo, Scott Roesch, Tobias Greitemeyer, and Bernard Weiner. 2004. "A meta-analytic review of help giving and aggression from an attributional perspective: Contributions to a general theory of motivation." *Cognition and Emotion* 18(6): 815-848.

Russell, Daniel W. 1982. "The causal dimension scale: A measure of how individuals perceive causes." *Journal of Personality and Social Psychology* 42(6): 1137-1145.

Safra, Zvi, Sinong Ma, and Tigran Melkonyan. 2019. "Is allocation affected by the perception of others' irresponsible behavior and by ambiguity?" *Risk Analysis* 39(10): 2182-2196.

Schmidt, Greg, and Bernard Weiner. 1988. "An attribution-affect-action theory of behavior: Replications of judgments of help-giving." *Personality and Social Psychology Bulletin* 14(3): 610-621.

Schokkaert, Erik, and Kurt Devooght. 2003. "Responsibility-sensitive fair compensation in different cultures." *Social Choice and Welfare* 21(2): 207-242.

Shonkwiler, J. Scott, and Nick Hanley. 2003. "A new approach to random utility modeling using the Dirichlet multinomial distribution." *Environmental and Resource Economics* 26(3): 401-416.

Shrout, Patrick E., and Joseph L. Fleiss. 1979. "Intraclass correlations: Uses in assessing rater reliability." *Psychological Bulletin* 86(2): 420-428.

Singer, Peter. 2009. *The Life You Can Save: Acting Now to End World Poverty*. Random House.

Small, Deborah A., and George Loewenstein. 2003. "Helping a victim or helping the victim: Altruism and identifiability." *Journal of Risk and Uncertainty* 26(1): 5-16.

Small, Deborah A., George Loewenstein, and Paul Slovic. 2007. "Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims." *Organizational Behavior and Human Decision Processes* 102(2): 143-153.

Trhal, Nadja, and Ralf Radermacher. 2009. "Bad luck vs. self-inflicted neediness – An experimental investigation of gift giving in a solidarity game." *Journal of Economic Psychology* 30(4): 517-526.

Tscharaktschiew, Nadine, and Udo Rudolph. 2016. "The who and whom of help giving: An attributional model integrating the help giver and the help recipient." *European Journal of Social Psychology* 46(1): 90-109.

Turmunkh, Uyanga, Martijn J. van den Assem, and Dennie van Dolder. 2019. "Malleable lies: Communication and cooperation in a high stakes TV game show." *Management Science* 65(10): 4795-4812.

van den Assem, Martijn J., Dennie van Dolder, and Richard H. Thaler. 2012. "Split or steal? Cooperative behavior when the stakes are large." *Management Science* 58(1): 2-20.

van Dolder, Dennie, Martijn J. van den Assem, Colin Camerer, and Richard H. Thaler. 2015. "Standing united or falling divided? High stakes bargaining in a TV game show." *American Economic Review* 105(5): 402-407.

Weiner, Bernard. 1974. *Achievement Motivation and Attribution Theory*. General Learning Press.

Weiner, Bernard. 1980a. "A cognitive (attribution)-emotion-action model of motivated behavior: An analysis of judgments of help-giving." *Journal of Personality and Social Psychology* 39(2): 186-200.

Weiner, Bernard. 1980b. "May I borrow your class notes? An attributional analysis of judgments of help giving in an achievement-related context." *Journal of Educational Psychology* 72(5): 676-681.

Weiner, Bernard. 1985. “An attributional theory of achievement motivation and emotion.” *Psychological Review* 92(4): 548-573.

Weiner, Bernard. 1986. *An Attributional Theory of Motivation and Emotion*. Springer-Verlag.

Weiner, Bernard. 1995. *Judgments of Responsibility: A Foundation for a Theory of Social Conduct*. Guilford Press.

Weiner, Bernard. 2006. *Social Motivation, Justice, and the Moral Emotions: An Attributional Approach.* Lawrence Erlbaum Associates.

West, Stephen G., and T. Jan Brown. 1975. “Physical attractiveness, the severity of the emergency and helping: A field experiment and interpersonal simulation.” *Journal of Experimental Social Psychology* 11(6): 531-538.

Yörük, Barış K. 2016. “Charity ratings.” *Journal of Economics & Management Strategy* 25(1): 195-219.

Yu, Ziyue, Shuai Yang, Hualu Zheng, and Xinyu Chang. 2024. “Charity ugliness premium: Donors’ empathy for unattractiveness during the COVID-19 pandemic and implications for charity advertising.” *Journal of Consumer Behaviour* 23(3): 1463-1477.

Appendix A: Example Episode

This appendix describes the three cases featured in Episode 46, and briefly discusses the donation outcomes of that episode in the light of our findings. The episode aired on July 6, 2015.

Candidate 1: Stephanie. Stephanie is 21 years old and studying to become a beautician. She was assigned male at birth but has identified as female since childhood. Her father disowned her when he found out she was wearing women's clothing, after which she ended up in youth care. At the time, she could not begin her gender transition because parental consent was required, and she had to wait until she turned 18. At the age of 18 she started hormone treatment, and a year later (now two years ago) she underwent gender reassignment surgery. However, the hormone treatment did not have the desired effect, and she still has virtually no breasts. Breast augmentation is considered a cosmetic procedure, and therefore not covered by her health insurance. Stephanie has no savings, because she had to take care of herself from a young age. She is requesting €3,000 for a breast augmentation. During the panel interview, Stephanie's determination is highlighted, and it is emphasized how important breast augmentation is to her and why she is unable to afford this herself.

Candidate 2: Sander. Sander is 39 years old and has a young family. Several years ago he had mental and behavioural issues and struggled with managing his finances, which led to debts accumulating to €20,000. He addressed the financial disorder by appointing a legal guardian. However, after completing three years of debt restructuring, it was discovered that thousands of euros in outstanding debts remained due to mistakes made by the legal guardian. He is requesting €2,600 to hire a lawyer to investigate what went wrong and hold the legal guardian accountable. During the panel interview, it is revealed that he has been diagnosed with attention deficit hyperactivity disorder (ADHD). Sander explains that he spent the €20,000 on rent, furniture and cars. The panel members note that he could also initiate a standard complaints procedure on his own and may not necessarily need a lawyer. They also express that he should have taken more action to monitor and control his own debts and payments.

Candidate 3: Mandy. Mandy is 35 years old and has a young family. She worked as a cleaner in the evenings. Her husband is a street paver. Although they never had much money, they were living a happy life. A year ago, Mandy was diagnosed with a highly aggressive bladder tumour. The tumour could not be completely removed, and has spread. She has been told she does not have much time left to live. She is requesting €4,500 to go on a family vacation. Due to financial constraints, they have not been able to afford a family vacation abroad for years, and such a vacation now is her last wish. The panel interview focuses on the physical discomfort she is currently experiencing, the country she wishes to visit with her family, and the important memories they could create together during the vacation.

Table A1: Summary statistics example episode

The table shows how much money the candidates of Episode 46 requested and received, along with raters' average assessments of their physical attractiveness (on a scale from 1 to 10), the well-being impact of giving them €100 or €10,000 (as a percentile score), and the four causal attributions (on nine-point Likert scales). Variables are defined as in Table 2.

	Stephanie	Sander	Mandy
Amount requested	3,000	2,600	4,500
Amount received	2,100	0	7,900
Attractiveness	7.00	4.25	6.50
Impact €100	31.3	18.5	54.5
Impact €10,000	82.5	56.8	90.5
Personal control	4.67	6.42	1.67
External control	3.75	6.17	1.58
Internal locus	7.67	4.92	7.58
Stability	5.25	2.83	6.92

Table A1 shows how much money the candidates requested and received, along with the averages of raters' assessments of their physical attractiveness, the well-being impacts of gifts of €100 and €10,000, and the four dimensions of causal attribution. Mandy's case scores relatively high in terms of donation impact and low on personal control, whereas the opposite holds for Sander. Stephanie's scores fall in between. As expected in light of our findings, the lion's share of the money was given to Mandy. The remainder was allocated to Stephanie, whereas Sander received nothing.

Appendix B: Instructions Rating Sessions

Dear participant,

This session in the BEELab is about giving behavior in a game show. After you have read these instructions, you will go to your cubicle. On the computer screen in your cubicle you will see a questionnaire. A total of nine videos from the Dutch TV show "Geld Maakt Gelukkig" are embedded in the questionnaire.

In each video clip, a candidate explains why he or she needs financial help. After each video clip, we will ask you questions about the candidate and the candidate's situation. We are interested in your honest opinion. Your answers are important for our research and will be stored anonymously.

Talking to other participants, using mobile phones or any other electronic devices is strictly prohibited. Mobile phones and other electronic devices should be left in the waiting room. Also, you are not allowed to use the computer in the BEELab for any other purpose than the questionnaire. If you are found violating these rules, you will forfeit any earnings from this session, and may be excluded from future sessions as well.

This session is finished when everybody is done with the survey, so there is no need to hurry. In total, the session is expected to last about 90 minutes. After everybody is done, the researcher in the BEELab will give you your payment of €13. By participating, you also automatically take part in a lottery of €100. We will contact the winner via e-mail after all sessions are finished.

After each video, we will ask you the following questions about the candidate and the candidate's situation:

Question 1a:

In general, gifts increase the well-being of the recipient and people who are indirectly involved. Assume that this impact can be measured, and that the overall impact of a gift is the sum of the increases in well-being for all people involved.

Imagine a group of 100 people, consisting of the candidate and 99 people who are randomly drawn from the Dutch population. Assume that people can be ranked in terms of the overall impact from receiving a monetary gift.

What would the rank of the candidate be if those 100 people are ordered from lowest overall impact (1) to highest overall impact (100) of a gift of €100?

Question 1b:

What would the rank of the candidate be if the 100 people are ordered from lowest overall impact (1) to highest overall impact (100) of a gift of €10,000? (Note that this rank can be lower than, higher than, or equal to your previous answer where the gift was €100.)

Question 2a:

Please give a brief description of the cause or the causes of the candidate's situation.

Question 2b:

Think about the cause or causes you have just described. The items below concern your impressions or opinions of this cause or these causes.

Is the cause, or are the causes, something…

that reflects an aspect of the candidate	9 - 1	that reflects an aspect of the situation
manageable by the candidate	9 - 1	not manageable by the candidate
permanent	9 - 1	temporary
the candidate can regulate	9 - 1	the candidate cannot regulate
over which others have control	9 - 1	over which others have no control
inside of the candidate	9 - 1	outside of the candidate
stable over time	9 - 1	variable over time
under the power of other people	9 - 1	not under the power of other people
about the candidate	9 - 1	about others
over which the candidate has power	9 - 1	over which the candidate has no power
unchangeable	9 - 1	changeable
other people can regulate	9 - 1	other people cannot regulate

Question 3:

What is your estimate of the candidate's age?

Question 4:

Please rate the candidate's physical appearance on a scale from 1 to 10, where 1 means "extremely unattractive" and 10 means "model beautiful or handsome".

In some video clips, two people ask for money. In those cases, please answer the questions about the candidate's age and the physical appearance for each person separately.

You can start the questionnaire after taking your seat in the cubicle. If you have any questions, please raise your hand and the researcher will come to you.

Appendix C: Correlation Matrices

Table C1: Correlation matrices

The table reports the correlation coefficients for the unstandardized (Panel A) and standardized variables (Panel B). Variables are defined as in Table 2. In Panel B, *Attractiveness*, *Impact €100*, *Impact €10,000*, *Personal control*, *External control*, *Internal locus*, and *Stability* are after the standardization of (underlying item) scores at the rater level; other variables are unstandardized. Asterisks denote statistical significance at the one ***, five **, and ten * percent level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Unstandardized												
(1) Amount requested	1.00											
(2) Amount received	0.18**	1.00										
(3) Age	-0.04	0.05	1.00									
(4) Female	0.06	0.15*	-0.11	1.00								
(5) White	-0.05	-0.17**	0.14*	-0.07	1.00							
(6) Attractiveness	0.05	0.05	-0.42***	0.19**	-0.20**	1.00						
(7) Impact €100	-0.25***	0.21***	0.03	0.15*	0.02	-0.08	1.00					
(8) Impact €10,000	0.19**	0.36***	0.17**	0.14*	-0.11	0.02	0.28***	1.00				
(9) Personal control	-0.20***	-0.39***	-0.13*	-0.21***	0.05	0.00	-0.10	-0.38***	1.00			
(10) External control	0.09	-0.02	0.16**	0.00	0.01	-0.15*	-0.01	0.08	0.01	1.00		
(11) Internal locus	-0.20***	-0.18**	-0.10	-0.13*	0.10	-0.09	0.02	-0.24***	0.16**	-0.44***	1.00	
(12) Stability	0.09	0.20**	0.17**	0.23***	0.03	0.00	0.10	0.10	-0.56***	-0.29***	0.16**	1.00
Panel B: Standardized												
(1) Amount requested	1.00											
(2) Amount received	0.18**	1.00										
(3) Age	-0.04	0.05	1.00									
(4) Female	0.06	0.15*	-0.11	1.00								
(5) White	-0.05	-0.17**	0.14*	-0.07	1.00							
(6) Attractiveness	0.06	0.04	-0.45***	0.19**	-0.14*	1.00						
(7) Impact €100	-0.21***	0.29***	0.04	0.14*	0.05	-0.18**	1.00					
(8) Impact €10,000	0.14*	0.48***	0.16**	0.29***	-0.12	0.03	0.37***	1.00				
(9) Personal control	-0.22***	-0.41***	-0.12	-0.26***	0.06	0.00	-0.26***	-0.43***	1.00			
(10) External control	0.09	-0.04	0.15*	-0.02	0.00	-0.09	-0.01	0.08	0.02	1.00		
(11) Internal locus	-0.21***	-0.20***	-0.12	-0.13*	0.09	-0.11	-0.01	-0.28***	0.22***	-0.51***	1.00	
(12) Stability	0.08	0.23***	0.17**	0.23***	0.01	-0.06	0.17**	0.22***	-0.58***	-0.32***	0.12	1.00

Appendix D: Histograms of Requested and Received Amounts

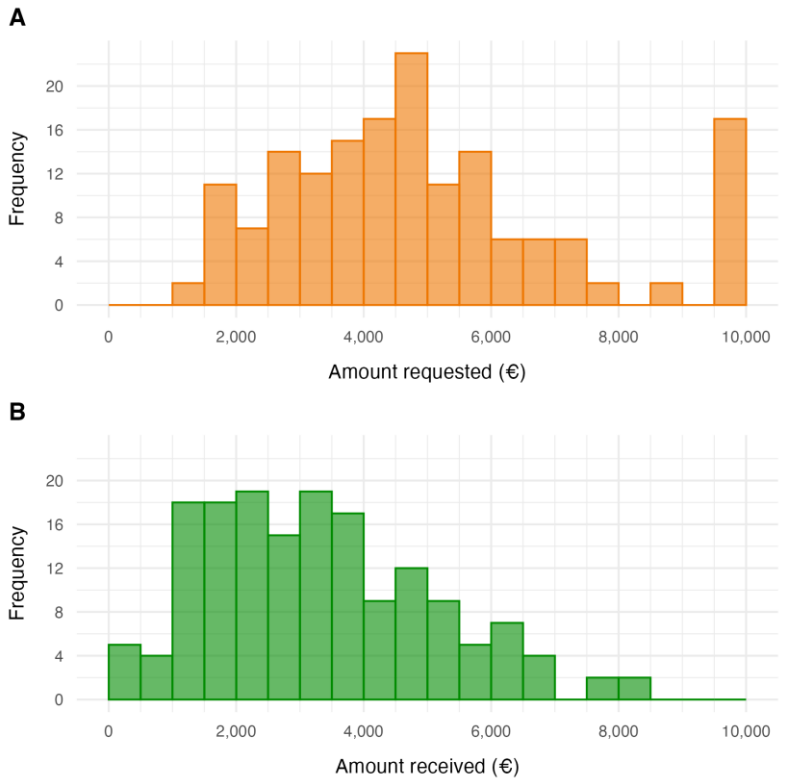

Figure D1: Histograms of requested and received amounts. This figure shows the distributions of the sums of money candidates asked for (Panel A) and received (Panel B), in euros. Bins are right-inclusive (with the lowest bin also including zero).

Appendix E: Ternary Plots Control Variables

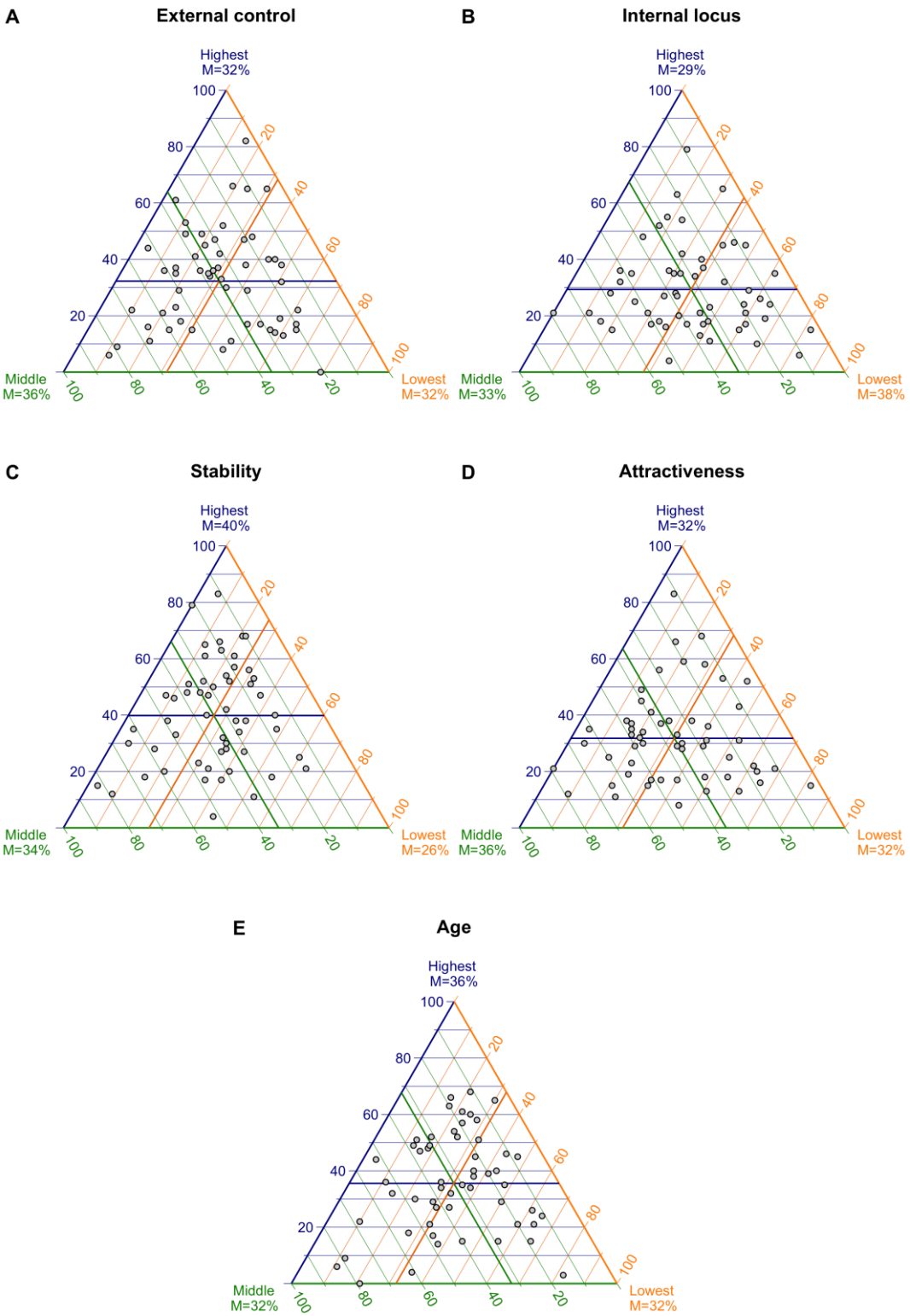

Figure E1: Ternary plots of donations in the show by rankings on *External control*, *Internal locus*, *Stability*, *Attractiveness*, and *Age*. The figure shows ternary plots for the percentage shares received by candidates ranking highest, middle, or lowest on *External control* (Panel A), *Internal locus* (B), *Stability* (C), *Attractiveness* (D), and *Age* (E). Variables are defined as in Table 2, and, with the exception of *Age*, after the standardization of (underlying item) scores at the rater level.