Seven (weak and strong) helping effects systematically tested in separate evaluation, joint evaluation and forced choice

Arvid Erlandsson*

Abstract

In ten studies (N = 9187), I systematically investigated the direction and size of seven helping effects (the identifiable-victim effect, proportion dominance effect, ingroup effect, existence effect, innocence effect, age effect and gender effect). All effects were tested in three decision modes (separate evaluation, joint evaluation and forced choice), and in their weak form (equal efficiency), or strong form (unequal efficiency). Participants read about one, or two, medical help projects and rated the attractiveness of and allocated resources to the project/projects, or choose which project to implement. The results show that the included help-situation attributes vary in their: (1) Evaluability – e.g., rescue proportion is the easiest to evaluate in separate evaluation. (2) Justifiability – e.g., people prefer to save fewer lives now rather than more lives in the future, but not fewer identified lives rather than more statistical lives. (3) Prominence – e.g., people express a preference to help females, but only when forced to choose.

Keywords: helping effects, moral decision making, separate and joint evaluation, forced choice, evaluability, justifiability, prominence, identified victim effect, proportion dominance, ingroup-bias.

1   Introduction

This article investigates the size and direction of seven helping effects. A "helping effect" occurs anytime a situational factor increases or decreases helping. Situational factors can

*Department of Behavioral Sciences and Learning, Linköping University, Sweden. Email: arvid.erlandsson@liu.se. https://orcid.org/0000-0001-7875-269X.

1113

be related to victim characteristics (e.g., Fisher & Ma, 2014), presentation of the problem (Erlandsson et al., 2018), surrounding context (Carroll et al., 2011), psychological distance (Ein-Gar & Levontin, 2013), type of request (Feiler et al., 2012) and more. Helping effects are frequently studied in social and organizational psychology, health economics and judgement and decision making, and typically investigated by experimentally manipulating a situational factor and then measuring prosocial motivation, behavioral intentions, or actual helping behavior.

This article examines seven well-known helping effects in a more rigorous and systematic way than what has been done previously. This is done by using a unified experimental paradigm, testing each helping effect in its "weak form" (implying equal efficiency) and in its "strong form" (implying unequal efficiency), as well as in three different decision modes (separate evaluation, joint evaluation and forced choice). This approach makes it possible to compare how the attributes central in the seven helping effects differ in evaluability, justifiability, and prominence (Li & Hsee, 2019; Slovic, 1975).

To facilitate understanding of key terms, I will first describe one of the seven helping effects, and use that effect to illustrate how "weak" and "strong" helping effects differ, and how helping effects can be tested in different decision modes. I will then describe the other six helping effects.

1.1   The identified victim effect

The identified victim effect (IVE) predicts that people are more motivated to help when the victims possible to save are identified than when they are non-identified (Butts et al., 2019; Kogut & Ritov, 2015; Lee & Feeley, 2016; Small & Loewenstein, 2003). Identifiability can be increased in several ways, such as adding a name, a photo, or a personal background story of the people in need (Kogut & Ritov, 2005a; Thornton et al., 1991). The underlying psychological mechanisms of the IVE are assumed to be different types of emotional reactions (e.g., distress, sympathy, anticipated warm glow if helping, or anticipated guilt if not helping) which are more intense when faced with identified victims than with non-identified statistical victims (Ahn et al., 2014; Erlandsson et al., 2015; Genevsky et al., 2013; Sah & Loewenstein, 2012). The IVE is an influential and widely investigated helping effect, but its robustness has been questioned (Hart et al., 2018; Lesner & Rasmussen, 2014; Perrault et al., 2015; Wiss et al., 2015), and the effect seems to come with several boundary conditions (Ein-Gar & Levontin, 2013; Friedrich & McGuire, 2010; Kogut, 2011; Kogut & Kogut, 2013; Kogut & Ritov, 2007; Smith et al., 2013).

Two studies in this article investigate the IVE. Study IVE1 takes place in a child cancer context whereas IVE2 takes place in a COVID-19 context. In both studies, helping motivation when faced with identified patients is compared against helping motivation when faced with non-identified patients.

1.2   The "weak" and "strong" forms of helping effects

Helping effects are sometimes seen as biases because they imply that lives are valued unequally (Baron & Szymanska, 2011; Caviola et al., 2020; Dickert et al., 2012; Slovic, 2007). A related distinction is the number of individuals possible to save in the opposing help projects. Sometimes, two projects which differ on only one attribute are compared against each other when testing a helping effect. At other times, the projects differ also on an efficiency-related attribute such as the number of victims possible to save, so that one option is superior on the efficiency attribute whereas the other is superior on the helping attribute of interest. I will refer to the former (equal efficiency) as a weak helping effect, and to the latter (unequal efficiency) as a strong helping effect. To exemplify, the "weak IVE" is here tested by comparing a project that can save 3 non-identified patients against a project that can save 3 identified patients, whereas the "strong IVE" is tested by comparing a project that can save 3 non-identified patients against a project that can save 1 identified patient.[1]

Importantly, to establish that a helping effect exists in its weak form does not imply that it exists also in its strong form. One could, e.g., find evidence for a weak IVE (i.e., people help more when they can help 3 identified compared to 3 non-identified), but at the same time find a reversed strong IVE (i.e., people help more when they can help 3 non-identified than when they can help 1 identified). With a few exceptions (Mata, 2016), past research has not clearly distinguished weak from strong helping effect.[2]

1.3   Testing helping effects in different decision modes: separate evaluation, joint evaluation and forced choice

Helping effects can be tested in different decision modes. Specifically, the distinction between separate evaluation and joint evaluation has been very influential for research on judgment and decision making (Bazerman et al., 1999; Bohnet et al., 2016; Hsee et al., 1999; Hsee et al., 2013; Hsee & Zhang, 2004; Paharia et al., 2009). In one famous study (Hsee, 1996), people had to express their willingness to pay for two dictionaries. Dictionary

---

[1]One often cited boundary condition that is related to the weak vs. strong IVE is that identifiability is said to boost helping only when the identified victim is presented alone (Small et al., 2007). Put differently, one identified victim elicits more helping than one statistical victim, but many identified victims does not elicit more helping than many statistical (Kogut & Ritov, 2005a). A related effect called the singularity effect predicts that people will help more when faced with a single identified victim than when faced with a group of identified victims (Dickert et al., 2016; Kogut & Ritov, 2005a; 2005b; Västfjäll et al., 2014). The two IVE-studies in this article are designed to test the weak IVE (3 identified vs. 3 non-identified patients) and the strong IVE (1 identified vs. 3 non-identified patients), but the experimental design will also allow us to test the singularity effect (1 identified vs. 3 identified).

[2]This paper focuses on the comparison between weak helping effects (where the projects can help equally many patients) and helping effects where one project can help exactly two more patients (e.g., 4 vs. 6). Strong helping effects can come in different magnitudes, so the results obtained here might not extrapolate to situations where the difference in the efficiency attribute is larger (e.g., 4 vs. 600) or smaller in proportion (e.g., 600 vs. 602).

A had 5000 entries and was in mint condition whereas dictionary B had 10000 entries but a torn cover. When one group of participants valued only A and another group valued only B (separate evaluation), it was found that mean willingness to pay was greater for dictionary A. Among participants who saw and valued both A and B (joint evaluation), the willingness to pay was instead greater for dictionary B, indicating an evaluation mode-elicited preference reversal. The suggested explanation for this is that the number of entries (but not the quality of the cover) is difficult to evaluate in separate evaluation, but that joint evaluation brings meaning to the number of entries by introducing a reference point. In line with this, emotional reactions predict attitudes toward policies more in separate evaluation (Ritov & Baron, 2011), whereas joint evaluation makes us more attentive to efficiency-related attributes (Bazerman et al., 2011; Caviola et al., 2014; Garinther et al., 2021).

Moreover, joint evaluation incorporates several decision modes (Fischer & Hawkins, 1993; Skedgel et al., 2015). When observing the options together, participants can, e.g., express preferences by rating the attractiveness or stating their willingness to pay for each alternative (joint-evaluation–rating), by distributing limited resources between the alternatives (joint-evaluation–allocation), or by choosing one of the alternatives (forced choice). A key difference is that ratings and allocations allow people to express indifference by rating the alternatives as equally attractive or by distributing resources evenly, whereas the choice mode forces people to favor one alternative (Sharps & Schroeder, 2019), even if this can be done randomly, e.g., by flipping a coin (Broome, 1984; Keren & Teigen, 2010). As predicted by the prominence effect (Slovic, 1975; Tversky et al., 1988), people do not choose randomly when faced with two alternatives that they previously rated as equally attractive, but instead tend to choose the alternative that is superior on the relatively more prominent attribute. This applies both when choosing which product to buy (Nowlis & Simonson, 1997), and when making moral choices about which people to help (Erlandsson et al., 2020).

This article will test both forms of the IVE in all three decision modes. The IVE in separate evaluation is tested by randomly assigning participants to read and respond to either a project that can save 3 identified patients, 1 identified patient or 3 non-identified patients. The weak IVE is tested by comparing attractiveness-ratings and allocations done by participants in the 3 identified condition against participants in the 3 non-identified condition. The strong IVE is tested by comparing those in the 1 identified condition against those in the 3 non-identified condition.

The IVE in joint evaluation is tested by having participants read about two helping projects presented side by side, rate the attractiveness of both projects, and allocate resources between the two projects. The weak [strong] IVE is tested by comparing ratings and allocations to the 3[1] identified-project against ratings and allocations to the 3 non-identified-project when the projects are presented together.

In the forced choice mode, participants read about two helping projects presented side by side as in joint evaluation, but are simply asked which of the two projects they prefer

1116

to implement. The weak [strong] IVE is found if significantly more than 50% of the participants choose the project that can save 3[1] identified when it is pitted against the project that can save 3 non-identified.

1.4   Underlying theory: Evaluable and justifiable attributes

Li and Hsee (2019) argue that attributes in decision situations can differ in both evaluability (how easy the attribute can be understood in itself) and in justifiability (whether people think the attribute should influence decisions). In a helping context, the number of people possible to help is a prime example of an attribute with high justifiability (as most people agree that it is preferable to help more than to help fewer people), but low evaluability (as it is difficult to assess if 3 patients are few or many without any comparison). As demonstrated in the above mentioned dictionary study, moving from separate to joint evaluation increases evaluability. For example, willingness to donate was no higher when one could save 200 rather than 100 polar bears in separate evaluation, but almost twice as high in joint evaluation (Hsee et al., 2013).

Other attributes in helping situations are different. For example, the identifiability attribute might have a moderately high evaluability (because identified beneficiaries tend to make us experience compassionate emotions even without comparison), but a relatively low justifiability (because most people do not believe that adding a name and a face should make a person more valuable). Applied to the IVE, the theory predicts that people will prefer to help fewer identified beneficiaries when the effect is assessed in separate evaluation (because identifiability is more evaluable), but that they will prefer to help more non-identified beneficiaries when assessed in joint evaluation (because efficiency is more justifiable). Consistent with this, a study by Kogut and Ritov (2005b, Study 2) found that participants reading about a project that could help one identified child donated more than participants reading about a project that could help a group of identified children (separate evaluation), but that the two projects received equal amounts when evaluated side by side (joint evaluation), and further that the project saving a group of children was preferred more often when participants were forced to choose one of the projects (i.e., a reversed effect). It should however be noted that this study manipulated singularity rather than identifiability.

The current article is the first to apply the theories suggested by Li and Hsee (2019) and by Slovic (1975) to a moral domain and expand previous research in at least two ways: (A) By systematically testing both weak and strong helping effects in both separate and joint evaluation, it will be possible to determine the evaluability and justifiability profile of a specific help-situation attribute, as well as how that attribute interact with the efficiency attribute. To exemplify, it is possible that people will prefer the identified project when the two projects are equally efficient (a weak IVE), but prefer the non-identified project when that is more efficient (a reversed strong IVE). (B) By testing joint evaluation preferences both with and without the option to express indifference, it is possible to detect choice-dependent moral preferences in line with the prominence theory. For example, people might prefer

the identified and non-identified projects equally often in ratings or in allocations, but still prefer the identified project more often when forced to choose.

Up until now, the IVE has been used to exemplify a helping effect. Importantly, this article includes no less than seven helping effects – all tested in a unified paradigm. This makes it possible to compare how various help-situation attributes (one in each helping effect) differ in their evaluability, justifiability, and prominence profiles. Below, I briefly describe the other included helping effects.

1.5 Six other helping effects

The article has been inspired by Kogut and Ritov (2005b) in regard to decision modes, and by Mata (2016) in regard to weak and strong helping effects. But whereas these studies investigated one helping effect each, I test seven effects.

1.5.1 The proportion dominance effect

The proportion dominance effect (PDE) predicts that people are more motivated to help when the rescue proportion is high (e.g., you can help 99% of the patients in need), than when it is low (e.g., you can help 1% of the patients in need; Baron, 1997; Bartels, 2006; Bartels & Burnett, 2011; Fetherstonhaugh et al., 1997; Friedrich et al., 1999; Jenni & Loewenstein, 1997; Kleber et al., 2013). Related research suggest that anticipated warm glow and helping responses towards one child possible to save is reduced when informing participants about other children that are not possible to help (Dickert & Slovic, 2009; Västfjäll et al., 2015), and that people are more motivated to help when they can save 100% of 100 victims in need than when they can save 100 victims without any denominator specified (Li & Chapman, 2009; Zhang & Slovic, 2018). The PDE is mediated by perceived impact, meaning that supporting a project with a low rescue proportion seems like a "drop in a bucket", whereas a high rescue proportion project seems more effective (Erlandsson et al., 2014, 2015).

In the two PDE-studies included in this article, the weak PDE is tested by comparing a project that can help 6 out of 6 patients in need against a project that can help 6 out of 100 patients in need. The strong PDE is tested by comparing a project that can help 4 out of 4 (in Study PDE1) or 4 out of 5 (in Study PDE2) against a project that can help 6 out of 100 (Mata, 2016).[3]

---

[3]Mata (2016) tested the weak and strong forms of the PDE in a forced choice mode and used the term "normative" for describing a weak effect and "non-normative" for describing a strong effect, thus suggesting that strong (but not weak) helping effects should be considered biases. I leave it to the readers to draw normative inferences, and therefore refer to "weak" and "strong" helping effects.

1.5.2   The ingroup effect

The ingroup effect (IGE, also known as parochialism or ingroup-bias) is a well-established phenomenon that predicts that people will help more when the people in need are from the helpers' ingroup than when they are from the helpers' outgroup (Baron, 2012; Duclos & Barasch, 2014; James & Zagefka, 2017; Fiedler et al., 2018; Levine & Thompson, 2004; Schwartz-Shea & Simmons, 1991). The IGE has been suggested to be driven by attitudes (e.g., ingroup-love and outgroup-hate; Brewer, 1999; De Dreu et al., 2011), beliefs (e.g., anticipated consequences for oneself; Everett et al., 2015), and by a greater perceived obligation and responsibility to help the ingroup (Erlandsson et al., 2015; Tomasello, 2020). Importantly, there are many types of ingroups, such as family, spatial proximity, shared values, or cultural identity, and different ways to classify ingroups can be perceived as separate helping effects (Waytz et al., 2019).

Two IGE-studies are included in this article. Study IGE1 focuses on kin-based ingroup (Burnstein et al., 1994), and tests the weak [strong] IGE by comparing a project that can help 3 relatives [1 relative] against a project that can help 3 unknown non-relatives. Study IGE2 focuses on nationality-based ingroup (Baron et al., 2013), and tests the weak [strong] IGE by comparing a project that can help 6[4] fellow citizens against a project that can help 6 foreigners.

1.5.3   The age effect

The age effect predicts that people will be more motivated to help when the people in need are young (children and teenagers) than when they are old (adults; Li et al., 2010). There are several possible reasons for this effect (Tsuchiya et al., 2003). One is that the evolved instinct to protect one's offspring can extrapolate to behavior towards children in general. Another is that children are perceived to be more dependent than adults, and unlike adults, young children are almost never held responsible for their own plight (Back & Lips, 1998). A third, more utilitarian reason for helping children is that the anticipated number of quality-adjusted life years is higher for a child than for an adult (Goodwin & Landy, 2014). Study AGE tests the weak [strong] age effect by comparing a project that can help 6[4] children and teenagers against a project that can help 6 adults.

1.5.4   The gender effect

The gender effect predicts that people will be more motivated to help when the people in need are women than when they are men (Dufwenberg & Muren, 2006; Eagly & Crowley, 1986; Weber et al., 2019). One explanation for this effect is that female participants are more motivated to help their gender-based ingroup, but males also seem to help women in need more than men in need. One reason for this is that helping by men can be used to signal affluence and agreeableness towards women (Raihani & Smith, 2015; van Vugt & Iredale, 2013). Another reason is that gender-stereotypes lead both men and women to

perceive females as less aggressive, more delicate, and more disadvantaged than men, and therefore both more deserving and in more need of protection (Bradley et al., 2019; Curry et al., 2004; Paolacci & Yalcin, 2020). Study GENDER tests the weak [strong] gender effect by comparing a project that can help 6[4] female patients against a project that can help 6 male patients.

1.5.5   The existence effect

The existence effect (aka the immediacy bias or the present bias; Cropper et al., 1994; Huber et al., 2011; O'Donoghue & Rabin, 2015) predicts that people are more motivated to help when it is possible to help individuals who are suffering now (existing victims) than when it is possible to help individuals who will suffer at some later point in time (future victims). The existence effect is much related to intertemporal choices and to the discounted utility model which suggests that utilities in the future are discounted by their delay (Bischoff & Hansen, 2016; Chapman & Elstein, 1995; Samuelson, 1937). In addition, the existence effect can be seen as the main psychological barrier for combatting climate-related threats as the primary beneficiaries of this type of helping are the future generations (Wade-Benzoni & Tost, 2009). Study EXISTENCE tests the weak [strong] existence effect by comparing a project that will start right away and help 6[4] existing patients, against a project that will start one year later and help 6 future patients.

1.5.6   The innocence effect

The innocence effect predicts that people are more motivated to help when it is possible to aid individuals who are the victims of unfortunate circumstances (external factors) than when it is possible to aid individuals who fully or partially caused their own plight, or who do not try to help themselves (internal factors; Fong, 2007; Lee et al., 2014; Seacat et al., 2007; Weiner, 1993). People report feeling less compassion and have less neural activity in areas associated with emotions when hearing about "non-innocent" victims (Fehse et al., 2015), and one study found that people suffering because of a natural disaster were helped more than people suffering from a civil war, due to a belief that natural disaster-victims try to help themselves more, and are less responsible for their current situation (Zagefka et al., 2011). Study INNOCENCE tests the weak [strong] innocence effect by comparing a project that can help 6[4] "innocent" patients who are ill despite exercising and eating healthy against a project that can help 6 "non-innocent" patients who smoke, drink, and eat excessively.

1.6   The current studies

There are multiple ways to test helping effects, and different methods, measures and contexts can create very diverging results. Rigorous and well-powered research that test different helping effects in a unified experimental paradigm is therefore much sought after. This

paper aims to do just this, as well as to test the size (and direction) of each helping effect in three decision modes and two forms. This research can help us understand the relative evaluability and justifiability (Li & Hsee, 2019) as well as the relative prominence (Slovic, 1975) of different helping effect attributes, explain past and future inconsistencies in the literature, and motivate researchers to take decision modes and the "weak" and "strong" forms into account when investigating helping effects and other types of moral decision making.

The seven included helping effects are among of the most frequently investigated in the prosocial decision making literature. The IVE, PDE and IGE were chosen in part because earlier research found that these effects are mediated by different psychological mechanisms (Erlandsson et al., 2015; 2017), suggesting that they might elicit different response patterns over the experimental manipulations. Still, the main contribution of this paper is not dependent on which helping effects are included, but rather that the included helping effects are tested much more systematically than what has been done before.

Three effects are tested in two studies each whereas the other four are tested in one preregistered study each. As all ten studies are similarly well-powered and adopt the same experimental design, use identical dependent variables, and have most contextual features in common, it will be possible to compare response patterns across the seven helping effects. If all helping effects are driven by the same underlying psychological mechanism, they would arguably be similarly affected when going from the weak to the strong form, and when moving between different decision modes.

This article could be said to investigate at least 42 research questions which can be derived from the following sentence: Does the [weak/strong] form of [IVE / PDE / IGE / AGE / GENDER / EXISTENCE / INNOCENCE] appear in the [separate evaluation / joint evaluation / forced choice] decision mode?

For each of these questions, the answer can be expressed with a percentage, where 100% indicates a very large helping effect (e.g., identified patients much favored over non-identified), 50% indicates absence of an effect, and 0% indicates a very large reversed helping effect (e.g., non-identified patients much favored over identified).

2   Method

All ten studies shared a similar core design and methodology. Participants were instructed to read and evaluate medical help projects, and randomly assigned to one out of seven conditions. Three conditions were used for testing the helping effects in separate evaluation, two were used for testing them in joint evaluation, and two were used for testing them in forced choice. I targeted 190–220 participants in each of the separate evaluation conditions

and 60–70 participants in each of the joint evaluation and forced choice conditions.[4] Please refer to all tables and to the online supplement for additional information about each study.[5]

2.1 Participants

Nine thousand one-hundred and eighty-seven complete responses were collected over ten studies (see Table 1). Data for the different studies were collected at different times but all participants were recruited from either Amazon Mechanical Turk or Prolific and payed $0.3–0.5.[6]

TABLE 1: Background information about each study.

Study name	Collection time (platform)	Females % (Mean age in years)	Total N	Valid N
PDE1	Spring19 (MTurk)	Not assessed	938	872
PDE2	Spring19 (MTurk)	Not assessed	861	778
IGE1	Summer20 (MTurk)	39.5% (35.35)	1108	863
IGE2	Spring19 (MTurk)	Not assessed	872	855
IVE1	Spring20 (Prolific)	73.7% (36.13)	862	845
IVE2*	Spring20 (Prolific)	54.3% (35.30)	1166	1135
Existence*	Fall19 (MTurk)	41.9% (37.91)	1005	951
Age*	Fall19 (MTurk)	44.1% (36.07)	977	935
Innocence*	Spring20 (MTurk)	34.9% (36.15)	1165	982
Gender*	Fall19 (MTurk)	45.4% (36.04)	1061	971

Note 1: Studies with "*" were preregistered.

Note 2: See the supplement for the number of participants in each experimental condition in each study.

[4]Sensitivity power analyses (using GPower; Faul et al., 2017) using the targeted cell n's and α = .05, power = 80% (two-tailed) determined the minimum detectable effect size to be d = 0.27–0.29 for SE, d = 0.34–0.37 for JE and g = 0.17–0.19 for CHOICE. Approximate sample sizes were based on both power analyses and financial constraints and decided before collecting data for each study. More participants were assigned to the SE-conditions to compensate for individual difference-related variance. Analyses were performed only after finishing each data collection.

[5]Find online supplement including raw data and preregistrations on https://osf.io/8fs46/?view_only=2f05b34b748642d08f645283e10062e4.

[6]Only American participants were recruited for eight of the studies whereas the two IVE-studies recruited participants from multiple English-speaking countries. The homogenous samples obviously pose a problem for generalizability, but this was done deliberately as the aim of this study was to maximize power and internal validity.

2.2   Material and procedure

2.2.1   Separate evaluation

Participants assigned to any of the separate evaluation conditions read and evaluated a single help project. Participants in Condition A(X) read about Project A which could treat a specified number of patients for a specified amount of money. Participants assigned to Condition A(X-2) read an identical description except that two fewer patients could be treated for the same amount of money. Seven of the ten conducted studies used the numbers "6" and "4" treated patients to operationalize "(X)" and "(X-2)" respectively. The other three studies (IGE2, IVE1 and IVE2) used the numbers "3" and "1". Participants assigned to Condition B(X) read about Project B which could treat equally many patients as A(X), but differed on one help-situation attribute which was different in the different studies and illustrated the helping effect currently being tested (see Tables 4–10).

Project A was presumed to be more attractive than Project B on the varying attribute in all studies, meaning that Project A could save: a higher proportion of patients in need (in PDE-studies), ingroup patients (in IGE-studies), identified patients (in IVE-studies), patients suffering now (EXISTENCE-study), children and teenagers (AGE-study), innocent "gymmers" (INNOCENCE-study) or female patients (GENDER-study; see the tables in the result section and the supplement).

The help project was presented to participants in a tabular form in eight studies (see Table 2 for an example and the supplement for all stimuli material). In the two IVE-studies, participants learned about the help project in written text rather than from a table (see the supplement).

TABLE 2: Example of how the help projects were presented to participants in separate evaluation (Study PDE1, condition A[6]). See the online supplement for all conditions in all studies.

	Project A
Who are affected by the disease?	People of all ages
In which country can the project be implemented?	USA (US patients will be treated)
Number of patients currently in need of treatment	6 patients currently need treatment
How effective is the treatment	The average chance of survival increases from 20% to 80% for patients that are treated
Number of patients that can be treated for $100,000	6 ill patients can be treated for $100,000 (100% of those in need)

Participants first responded to three attention check questions, meaning that they repeated provided information about the project. Participants who could not do this were deemed

inattentive and screened out (see Table 1).[7]

Next, participants were asked to rate the attractiveness of the helping project based on the provided information by responding to three questions; "how good does Project A[B] seem to you", "how worthy of financing does Project A[B] seem to you" and "how much do you approve of implementing Project A[B]". Participants responded on a visual analog scale ranging from 0 (not at all) to 100 (extremely) without any additional labels. Participants could see the number for where the marker was currently placed. These three questions were aggregated into a single variable labeled "rating" (all α's > .80).[8]

Thereafter, participants were asked to state how much of a hypothetical budget they wanted to earmark to the described project and to unspecified "other projects" respectively. In order to anchor participants' responses, participants were told that "the default allocation for a help project is 20%" but that they could earmark more if they found the project specifically worthy of financing. The percentage they earmarked to the project at hand (0–100%, same type of scale as for ratings) was labeled "allocation".[9]

The weak helping effects (equal efficiency) were tested by comparing the rating- and allocation-scores of participants reading about Project A(X) against those reading about Project B(X), whereas the strong helping effects (unequal efficiency) were tested by comparing ratings and allocations of those reading A(X-2) against those reading B(X).

2.2.2   Joint evaluation

Participants assigned to the joint evaluation conditions, read about two help projects presented next to each other, and evaluated both projects. In eight of the studies, this was done by adding a column in the tables so that participants could easily compare the two projects on each attribute (see Table 3). In the two IVE-studies, an additional paragraph of text described the second help project (see the supplement).

Half of the participants in joint evaluation read about Project A(X) and Project B(X) presented side by side (testing the weak effect), whereas the other half read about Project A(X-2) and Project B(X) presented side by side (testing the strong effect).

The attention check questions used in separate evaluation were used in joint evaluation as well, with the only difference that participants had to respond to questions regarding both Project A and Project B. The three questions used to assess attractiveness ratings were used

---

[7]Only two attention check questions were used in Studies IGE2, IVE1 and IVE2.

[8]Kahneman et al., (1993), argue that willingness to donate (like other contingent valuations) are measuring basically the same construct as attitude scales, which also tend to have superior psychometrical properties. Therefore, I opted to measure attractiveness/importance of helping projects rather than with willingness to donate.

[9]After responding to all questions about the first help project, participants in the separate evaluation conditions read about another help project (B[X], B[X] and A[X] for A[X], A[X-2] and B[X] respectively), and responded to the same questions. Results from these questions are not included in the manuscript as they represent a hybrid of separate and joint evaluation, but they are reported in the supplement and included in the raw data files.

TABLE 3: Example of how the help projects were presented to participants in joint evaluation and forced choice (Study PDE1, condition A[4] vs. B[6]). See the online supplement for all conditions in all studies.

	Project A	Project B
Who are affected by the disease?	People of all ages	People of all ages
In which country can the project be implemented?	USA (US patients will be treated)	USA (US patients will be treated)
Number of patients currently in need of treatment	4 patients currently need treatment	100 patients currently need treatment
How effective is the treatment	The average chance of survival increases from 20% to 80% for patients that are treated	The average chance of survival increases from 20% to 80% for patients that are treated
Number of patients that can be treated for $100,000	4 ill patients can be treated for $100,000 (100% of those in need)	6 ill patients can be treated for $100,000 (6% of those in need)

also in joint evaluation (all α's > .80). Participants first responded to the three questions regarding Project A and then to the same three questions regarding Project B.

The allocation task in joint evaluation was different from the one used in separate evaluation. Participants were asked to allocate resources only between Projects A and B and explicitly told to allocate 50–50 in case they found both projects equally worthy of financing.

2.2.3 Forced choice

In the forced choice-conditions, participants read the same information and responded to the same attention check questions as in the joint evaluation-conditions. Half of the participants read A(X) vs. B(X) for testing the weak effect, the other half read A(X-2) vs. B(X) for testing the strong effect. However, rather than evaluating the projects using ratings and allocations, participants had to choose which of the two projects to implement. Participants could not refrain from choosing, but those who found the projects equally attractive were suggested to use an embedded online number generator to guide their choice (see the supplement). The number of participants who used the number generator was not recorded.

3   Results

The results are organized so that the seven helping effects are presented one at the time, beginning with a short summary of the results. The weak form (when the two projects can treat equally many patients) and the strong form (when Project A — presumed to be more attractive on the varying attribute — can help fewer patients) are presented separately for each effect.

The weak and strong forms of all helping effects were tested in separate evaluation (independent-sample t-test), joint evaluation (paired t-test) and with forced choice (one proportion binomial test).[10]   Tables 4–10 (one table per helping effect) show cell means for ratings and allocations, the number of participants choosing each project, and the corresponding statistical test (unadjusted p-values).

Beyond testing the size and direction of each helping effect, I also aimed to compare the effects in two ways: (1) separate vs. joint evaluation; (2) preferences expressed in joint evaluation vs. forced choices.

The first comparison is complicated by the fact that effect sizes from between-group comparisons are not easily comparable with effect sizes from within-subject comparisons, because the unavoidable additional variance when comparing different subjects. I therefore compared mean differences instead. Specifically, for both separate and joint evaluation comparisons, I calculated a Project A minus Project B mean difference score for ratings and allocations (both measured on 0–100 scales). A positive mean difference score illustrates a helping effect, a score around zero indicates absence of an effect, and a negative mean difference score indicates a reversed helping effect. I then compared mean difference scores obtained in separate and joint evaluation. A higher[lower] mean difference score in joint evaluation indicates that joint evaluation increases[reduces] the helping effect.

For the second comparison, I calculated the percentage of participants (in joint evaluation-conditions) who expressed a preference for Project A by rating it higher or by allocating more than 50% of the resources to it (see Tables 4–10). Participants who gave equal ratings or allocated 50–50 were split so that exactly half of them preferred each project (when an uneven number of participants gave equal ratings or allocations, one was excluded). These rating- and allocation-inferred preferences were then compared against the preferences expressed in forced choice with 2*2 chi-square tests.

The percentage scores in the rightmost column in Tables 4–10 denotes different things for different rows. The percentage for separate (SE) and joint evaluation (JE) ratings and allocations is the "common language effect size" for each comparison of means (Lakens, 2013; McGraw & Wong, 1992). For independent t-tests (separate evaluation), the percentage expresses the probability that that a randomly sampled person reading about Project A (the project presumed to be more attractive on the varying attribute) have a higher observed value than a randomly sampled individual reading about Project B. For paired t-tests (joint

---

[10]Very similar results were found using equivalent non-parametric tests (Mann-Whitney U and Wilcoxon), see the supplement.

evaluation), the percentage indicates the likelihood that a randomly selected person rates Project A higher than Project B (Lakens, 2013).[11] The percentage score for rows labeled "preferences" denotes the proportion of participants who preferred Project A over Project B when they were forced to choose, and when preferences were inferred from ratings and allocations. A high percentage (green cells) indicates presence of a helping effect, a low percentage (orange cells) indicates presence of a reversed helping effect, and a percentage around 50% (yellow cells) indicates absence of any effect.

3.1 Proportion dominance effect (Studies PDE1 and PDE2)

The weak PDE was found in all three decision modes and not consistently affected by decision modes. The strong PDE was clearly present in separate evaluation, weaker in joint evaluation, and weaker still in forced choice.

3.1.1 Weak PDE (6 out of 6 patients vs. 6 out of 100 patients)

Separate evaluation.   Participants reading about a high rescue proportion project gave higher attractiveness ratings than participants reading about a low rescue proportion project helping equally many (M = 79.25 vs. 44.24 in PDE1 and 78.38 vs. 49.92 in PDE2). Those reading about the high proportion project also earmarked more resources (M = 51.25 vs. 28.30 in PDE1 and 49.24 vs. 33.70 in PDE2).

Joint evaluation.   When presented side by side, participants rated the high proportion project as more attractive than the low proportion project helping equally many (M = 80.17 vs. 57.28 in PDE1 and 83.05 vs. 54.09 in PDE2). They also allocated more resources to the high proportion project (M = 68.17 vs. 31.83 in PDE1 and 66.30 vs. 33.70 in PDE2).

The mean difference (Project A [6 of 6] minus Project B [6 of 100]) for ratings was around 12 points higher in separate than in joint evaluation (SE = 35.01, JE = 22.89) in PDE1, and about the same in PDE2 (SE = 28.46, JE = 28.96). In contrast, the mean difference in allocations was lower in separate evaluation in both studies (SE = 22.95, JE = 36.34 in PDE1; SE = 15.54, JE = 32.6 in PDE2).

Forced choice.   83.82% in PDE1, and 76.79% in PDE2 chose to implement the high proportion project when the two projects helped equally many patients. When aggregating both PDE-studies, it was found that preferences expressed with forced choice did not differ from preferences inferred from joint evaluation ratings ($\chi 2$ = 1.47, p = .226) or allocations ($\chi 2$ = 1.13, p = .288).[12]

---

[11]I used the spreadsheet created by Lakens (2013) to calculate the common language effect size (input and output values can be found in the supplement). For reasons explained above, the common language effect sizes from separate and joint evaluation cannot be directly compared against each other.

[12]See the supplement for results for each study separately.

TABLE 4: Results for the proportion dominance effect studies (PDE1 and PDE2).

	Project A	Project B	Equal	Test	Percentage
Weak effect	**A(X)**	**B(X)**			
PDE1	**6 of 6**	**6 of 100**			
SE rating	79.25 (20.71)	44.24 (27.59)		t[389] = 14.20, p < .001	84.49%
SE allocation	51.25 (30.58)	28.30 (19.06)		t[389] = 8.90, p < .001	73.79%
JE rating	80.17(18.55)	57.28 (24.83)		t[63] = 5.86, p < .001	76.79%
JE allocation	68.17 (22.74)	31.83 (22.74)		t[63] = 6.39, p < .001	78.79%
JE rating preference	44	14	6	z = 3.75, p < .001	73.44%
JE allocation preference	43	9	12	z = 4.25, p < .001	76.56%
Forced choice preference	57	11		z = 5.58, p < .001	83.82%
PDE2	**6 of 6**	**6 of 100**			
SE rating	78.38 (21.04)	49.92 (25.95)		t[348] = 11.26, p <.001	80.29%
SE allocation	49.24 (27.81)	33.70 (22.19)		t[348] = 5.78, p <.001	66.89%
JE rating	83.05 (20.37)	54.09 (28.99)		t[56] = 6.66, p <.001	81.11%
JE allocation	66.30 (26.80)	33.70 (26.80)		t[56] = 4.59, p <.001	72.85%
JE rating preference	38	10	9	z = 3.71, p < .001	74.56%
JE allocation preference	35	9	13	z = 3.44, p < .001	72.81%
Forced choice preference	43	13		z = 4.01, p < .001	76.79%
Strong effect	**A(X-2)**	**B(X)**			
PDE1	**4 of 4**	**6 of 100**			
SE rating	73.37 (26.42)	44.24 (27.59)		t[395] = 10.75, p < .001	77.71%
SE allocation	43.67 (29.26)	28.30 (19.06)		t[395] = 6.17, p < .001	67.01%
JE rating	79.88 (19.23)	63.44 (27.28)		t[74] = 4.19, p <.001	68.57%
JE allocation	54.47 (28.12)	45.53 (28.12)		t[74] = 1.38, p = .173	56.32%
JE rating preference	43	27	5	z = 1.85, p = .065	60.67%
JE allocation preference	37	29	9	z = 0.92, p = .356	55.33%
Forced choice preference	33	38		z = −0.59, p = .553	46.48%
PDE2	**4 of 5**	**6 of 100**			
SE rating	73.83 (20.50)	49.92 (25.95)		t[369] = 9.89, p <.001	76.52%
SE allocation	46.38 (28.11)	33.70 (22.19)		t[369] = 4.79, p <.001	63.84%
JE rating	70.01 (23.40)	57.53 (30.08)		t[61] = 2.72, p = .009	63.52%
JE allocation	50.74 (29.37)	49.26 (29.37)		t[61] = 0.20, p = .843	51.01%
JE rating preference	32	24	6	z = 1.02, p = .301	56.45%
JE allocation preference	27	28	7	z = −0.13, p = .899	49.19%
Forced choice preference	22	36		z = −1.84, p =.066	38.97%

1128

3.1.2   Strong PDE (4 out of 4[5] patients vs. 6 out of 100 patients)

Separate evaluation.   Participants reading about a high rescue proportion project treating four patients, gave higher attractiveness ratings than participants reading about a low rescue proportion project treating six patients (M = 73.37 vs 44.24 in PDE1 and 73.83 vs. 49.92 in PDE2). Those reading about the high proportion project also earmarked more resources (M = 43.67 vs. 28.30 in PDE1 and 46.38 vs. 33.70 in PDE2).

Joint evaluation.   When presented side by side, participants rated the high proportion project treating four as more attractive than the low proportion project treating six patients (M = 79.88 vs. 63.44 in PDE1 and 70.01 vs. 57.53 in PDE2). However, they allocated resources about evenly (M = 54.47 vs. 45.43 in PDE1 and 50.74 vs. 49.26 in PDE2).

The mean difference (Project A [4 of 4 or 4 of 5] minus Project B [6 of 100]) for ratings was around 12 points higher in separate than in joint evaluation in both PDE1 (SE = 29.13, JE = 16.44) and PDE2 (SE = 23.91, JE = 12.48). Likewise, the mean difference for allocations was higher in separate evaluation in both PDE1 (SE = 15.37, JE = 8.94) and PDE2 (SE = 12.68, JE = 1.48). This indicates that joint evaluation slightly reduces the strong PDE.

Forced choice.   46.48% (in PDE1), and 37.93% (in PDE2) chose to implement the high proportion project when that project helped fewer patients. When aggregating both PDE-studies, it was found that participants were slightly less likely to express preferences in line with the strong PDE in forced choice than in attractiveness ratings ($\chi 2$ = 6.94, p = .008), but not than in resource allocations ($\chi 2$ = 2.62, p = .105).

3.2   Ingroup effect (Studies IGE1 and IGE2)

The weak IGE was found in all three decision modes when the salient ingroup was family [IGE1], but only in joint evaluation allocations and in forced choice when the salient ingroup was nationality [IGE2]. Joint evaluation (for allocations) increased the weak IGE, and forced choice increased it further. The strong IGE was not found in any decision mode when the ingroup was fellow citizens [IGE2] but it was found in separate evaluation and forced choice when the ingroup was kin [IGE1]. Joint evaluation did not consistently affect the strong IGE, but expressing preferences with forced choice increased it.

3.2.1   Weak IGE (3[6] ingroup patients vs. 3[6] outgroup patients)

Separate evaluation.   Participants reading about a project treating relatives gave higher attractiveness ratings (M = 72.54 vs. 63.42) and also earmarked more resources (M = 52.68 vs. 41.22) than those reading about a project treating equally many non-relatives in IGE1. In contrast, participants reading about a project treating fellow citizens gave similar ratings (M

TABLE 5: Results for the ingroup effect studies (IGE1 and IGE2).

	Project A	Project B	Equal	Test	Percentage
Weak effect	**A(X)**	**B(X)**			
IGE1	**3 relatives**	**3 unknown**			
SE rating	72.54 (23.07)	63.42 (24.65)		$t[374] = 3.71$, $p < .001$	60.65%
SE allocation	52.68 (25.66)	41.22 (23.50)		$t[374] = 4.50$, $p < .001$	62.91%
JE rating	79.59 (22.26)	71.88 (24.09)		$t[72] = 2.96$, $p = .004$	63.53%
JE allocation	62.66 (19.11)	37.34 (19.11)		$t[72] = 5.66$, $p < .001$	74.62%
JE rating preference	40	16	17	$z = 2.81$, $p = .005$	66.44%
JE allocation preference	41	7	25	$z = 3.98$, $p < .001$	73.29%
Forced choice preference	60	10		$z = 5.98$, $p < .001$	85.71%
IGE2	**6 US**	**6 Polish**			
SE rating	69.79 (24.87)	71.51 (25.60)		$t[391] = -0.68$, $p = .499$	48.08%
SE allocation	38.24 (23.28)	40.14 (23.15)		$t[391] = -0.81$, $p = .418$	47.69%
JE rating	76.04 (20.98)	72.83 (24.08)		$t[64] = 1.58$, $p = .120$	57.73%
JE allocation	56.89 (15.03)	43.11 (15.03)		$t[64] = 3.70$, $p < .001$	67.67%
JE rating preference	19	20	26	$z = -0.12$, $p = .901$	49.23%
JE allocation preference	18	1	46	$z = 2.11$, $p = .035$	63.08%
Forced choice preference	57	10		$z = 5.74$, $p < .001$	85.07%
Strong effect	**A(X-2)**	**B(X)**			
IGE1	**1 relative**	**3 unknown**			
SE rating	70.25 (25.18)	63.42 (24.65)		$t[374] = 2.65$, $p = .008$	57.68%
SE allocation	50.14 (25.75)	41.22 (23.50)		$t[374] = 3.49$, $p = .001$	60.10%
JE rating	70.62 (19.46)	75.57 (20.52)		$t[71] = -2.03$, $p = .046$	40.54%
JE allocation	53.83 (23.39)	46.17 (23.39)		$t[71] = 1.39$, $p = .169$	56.50%
JE rating preference	24	42	6	$z = -2.12$, $p = .034$	37.50%
JE allocation preference	31	27	14	$z = 0.47$, $p = .637$	52.78%
Forced choice preference	52	22		$z = 3.49$, $p < .001$	70.23%
IGE2	**4 US**	**6 Polish**			
SE rating	68.40 (25.64)	71.51 (25.60)		$t[392] = -1.21$, $p = .229$	46.58%
SE allocation	42.27 (27.50)	40.14 (23.15)		$t[392] = 0.83$, $p = .407$	52.36%
JE rating	70.82 (19.44)	75.26 (19.96)		$t[65] = -4.16$, $p < .001$	30.38%
JE allocation	49.47 (19.42)	50.53 (19.42)		$t[65] = -0.22$, $p = .825$	48.91%
JE rating preference	14	43	9	$z = -3.57$, $p < .001$	28.03%
JE allocation preference	15	33	18	$z = -2.22$, $p = .027$	36.36%
Forced choice preference	30	34		$z = -0.50$, $p = .618$	46.88%

1130

= 69.79 vs. 71.51) and earmarked similar amounts (M = 38.24 vs. 40.14) as those reading about a project treating equally many foreigners in IGE2.

Joint evaluation.   When evaluated side by side, the project treating relatives was rated as more attractive than the project treating equally many non-relatives (M = 79.59 vs. 71.88), and also allocated more resources in IGE1 (M = 62.66 vs. 37.34). The project treating fellow citizens was rated as non-significantly more attractive than the project treating equally many foreigners (M = 76.04 vs. 72.83), and also allocated more resources in IGE2 (M = 56.89 vs. 43.11).

The mean difference (Project A [X ingroup patients] minus Project B [X outgroup patients]) in ratings was similar for separate and joint evaluations in both IGE1 (SE = 9.12, JE = 7.71) and in IGE2 (SE = −1.72, JE = 3.21). However, the mean difference in allocations was higher in joint evaluation in both studies (SE = 11.46, JE = 25.32 in IGE1; SE = -1.90, JE = 13.78 in IGE2).

Forced choice.   85.71% in IGE1 (kin) and 85.07% in IGE2 (nationality) chose to help ingroup rather than outgroup patients when the two projects helped equally many patients. When aggregating both IGE-studies, it was found that participants were more likely to express preferences in line with the weak IGE in forced choice, than in attractiveness ratings ($\chi2$ = 24.73, p < .001) or in resource allocations ($\chi2$ = 10.90, p < .001).

3.2.2   Strong IGE (1[4] ingroup patients vs. 3[6] outgroup patients)

Separate evaluation.   Participants reading about a project treating one relative gave slightly higher attractiveness ratings than those reading about a project treating three unknown patients (M = 70.25 vs. 63.42), and they also earmarked more resources in IGE1 (M = 50.14 vs. 41.22). In contrast, participants reading about a project treating four fellow citizens gave similar ratings as those reading about a project treating six foreigners (M = 68.40 vs. 71.51), and they also earmarked similar amounts of resources in IGE2 (M = 42.27 vs. 40.14).

Joint evaluation.   When evaluated side by side, the project treating more outgroup patients was rated as more attractive than the project treating fewer ingroup patients in both studies (M = 70.62 vs. 75.57 in IGE1 and 70.82 vs. 75.26 in IGE2). The two project-pairs were however allocated equal amounts of resources (M = 53.83 vs. 46.17 in IGE1 and 49.47 vs. 50.53 in IGE2).

The mean difference (Project A [fewer ingroup patients] minus Project B [more outgroup patients]) in ratings was around 12 points higher in separate evaluation when the ingroup was kin in IGE1 (SE = 6.83, JE = -4.95), but about the same when the ingroup was nationality in IGE2 (SE = −3.11, JE = −1.06). The mean difference in allocations was similar in both studies (SE = 8.92, JE = 7.66 in IGE1; SE = 2.13, JE = −1.06 in IGE2).

Forced choice.   70.27% chose to treat one relative rather than three unknown patients (in IGE1), whereas only 46.88% chose to treat four fellow citizens rather than six foreigners (in IGE2). When aggregating both IGE-studies, it was found that participants were more likely to express preferences in line with the strong IGE in forced choice, than in attractiveness ratings ($\chi2 = 19.53$, p < .001), or in resource allocations ($\chi2 = 5.81$, p = .016).

3.3   Identified victim effect (Studies IVE1 and IVE2)[13]

The weak IVE was found to some extent in all three decision modes. On the contrary, No strong IVE was found in any decision mode. Instead, participants expressed clear preferences for saving a greater number of non-identified victims in joint evaluation and forced choice (i.e., a reversed strong IVE).

3.3.1   Weak IVE (3 identified patients vs. 3 non-identified patients)

Separate evaluation.   Participants reading about a project treating identified patients gave higher attractiveness ratings than those reading about a project treating equally many non-identified patients (M = 87.25 vs. 79.53 in IVE1 and 70.10 vs. 60.18 in IVE2). Still, the two groups earmarked similar amounts of resources (M = 59.12 vs. 54.65 in IVE1 and 32.71 vs. 33.11 in IVE2).

Joint evaluation.   When presented side by side, the identified patients-project was rated as slightly more attractive than the non-identified patient-project helping equally many (89.27 vs. 84.49 in IVE1 and 77.48 vs. 72.49 in IVE2). The project helping identified patients was also allocated more resources in both studies (M = 55.86 vs. 44.14 in IVE1 and 52.84 vs. 47.16 in IVE2).

The mean difference (Project A [3 identified] minus Project B [3 non-identified]) in ratings was slightly higher in separate evaluation (SE = 7.72, JE = 4.78 in IVE1; SE = 9.92, JE = 4.99 in IVE2). On the contrary, the mean difference in allocations was slightly higher in joint evaluation (SE = 4.47, JE = 11.72 in IVE1; SE = −0.40, JE = 5.68 in IVE2).

Forced choice.   72.58% in IVE1 and 64.84% in IVE2 chose to implement the project helping three identified patients rather than the project helping three non-identified. When aggregating both IVE-studies, it was found that preferences expressed with forced choice did not differ from preferences inferred from joint evaluation ratings ($\chi2 = 0.56$, p = .454) or allocations ($\chi2 = 1.36$, p = .244).

---

[13]Study IVE2 balanced the presentation order of the identified and non-identified project as well as the identity of the single identified patient. Additional descriptive data for this study are reported in the supplement.

TABLE 6: Results for the identified victim studies (IVE1 and IVE2).

	Project A	Project B	Equal	Test	Percentage
Weak effect	**A(X)**	**B(X)**			
IVE1	**3 identified**	**3 statistical**			
SE rating	87.25 (14.56)	79.53 (19.54)		t[385] = 4.41, p <.001	62.43%
SE allocation	59.12 (29.50)	54.65 (28.68)		t[385] = 1.51, p = .132	54.33%
JE rating	89.27 (12.31)	84.49 (17.69)		t[63] = 3.49, p <.001	66.87%
JE allocation	55.86 (12.46)	44.14 (12.46)		t[63] = 3.76, p <.001	68.09%
JE rating preference	24	11	29	z = 1.63, p = .104	60.16%
JE allocation preference	19	0	45	z = 2.37, p = .018	64.84%
Forced choice preference	45	17		z = 3.56, p <.001	72.58%
IVE2	**3 identified**	**3 statistical**			
SE rating	70.10 (23.28)	60.18 (27.35)		t[445] = 4.13, p <.001	60.88%
SE allocation	32.71 (19.50)	33.11 (21.46)		t[445] = −0.21, p = .836	49.45%
JE rating	77.48 (20.20)	72.49 (23.70)		t[88] = 2.83, p =.006	61.79%
JE allocation	52.84 (10.37)	47.16 (10.37)		t[88] = 2.59, p =.011	60.79%
JE rating preference	37	21	31	z = 1.70, p = .090	58.99%
JE allocation preference	23	7	59	z = 1.70, p = .090	58.99%
Forced choice preference	59	32		z = 2.83, p = .005	64.84%
Strong effect	**A(X-2)**	**B(X)**			
IVE1	**1 identified**	**3 statistical**			
SE rating	80.12 (19.07)	79.53 (19.54)		t[389] = 0.30, p = .764	50.86%
SE allocation	56.58 (28.98)	54.65 (28.68)		t[389] = 0.66, p = .510	51.89%
JE rating	63.35 (23.18)	86.77 (13.58)		t[64] = −8.11, p < .001	15.73%
JE allocation	30.98 (14.96)	69.02 (14.96)		t[64] = −10.25, p < .001	10.18%
JE rating preference	5	55	5	z = −6.20, p < .001	11.54%
JE allocation preference	3	53	9	z = −6.20, p < .001	11.54%
Forced choice preference	9	59		z = −6.06, p < .001	13.24%
IVE2	**1 identified**	**3 statistical**			
SE rating	56.21 (28.05)	60.18 (27.35)		t[471] = −1.55, p = .121	45.96%
SE allocation	27.93 (18.98)	33.11 (21.46)		t[471] = −2.79, p = .006	42.83%
JE rating	49.15 (27.50)	76.82 (19.39)		t[128] = −12.56, p < .001	13.43%
JE allocation	28.26 (15.39)	71.74 (15.39)		t[128] = −16.04, p < .001	7.89%
JE rating preference	7	115	7	z = −9.51, p < .001	8.14%
JE allocation preference	4	112	13	z = −9.51, p < .001	8.14%
Forced choice preference	17	113		z = −8.42, p < .001	13.08%

1133

3.3.2   Strong IVE (1 identified patient vs. 3 non-identified patients)

Separate evaluation.   Participants reading about a project treating one identified patient gave similar attractiveness ratings as those reading about a project treating three non-identified patients (M = 80.12 vs. 79.53 in IVE1 and 56.21 vs. 60.18 in IVE2). Earmarked resources were similar in the two groups when the patients were children in IVE1 (M = 56.58 vs. 54.65), but participants reading about a project treating three non-identified earmarked slightly more than those reading about one identified when the patients were adults in IVE2 (M = 27.93 vs. 33.11).[14]

Joint evaluation.   When presented side by side, the project helping three non-identified patients was rated as much more attractive than the project helping one identified (M = 63.35 vs. 86.77 in IVE1 and 49.15 vs. 76.82 in IVE2). It was also allocated a larger portion of the resources (M = 30.98 vs. 69.02 in IVE1 and 28.26 vs. 71.74 in IVE2).

The mean difference (Project A [1 identified] minus Project B [3 non-identified]) in ratings was around 24 points lower (more negative) in joint evaluation in both studies (SE = 0.59, JE = −23.42 in IVE1; SE = -3.97, JE = −27.67 in IVE2). Likewise, the mean difference in allocations was much lower in joint evaluation (SE = 1.93, JE = −38.04 in IVE1; SE = -5.18, JE = −43.48 in IVE2). This clearly indicates that joint evaluation reverses the strong IVE.

Forced choice.   Only 13.24% (in IVE1) and 13.08% (in IVE2) chose to implement the project that could treat one identified over the project that could treat three non-identified patients. When aggregating both IVE-studies, it was found that preferences expressed with forced choice did not differ from preferences inferred from joint evaluation ratings or allocations (both $\chi 2$ = 1.82, p = .178).

3.4   Existence effect

Both the weak and the strong form of the existence effect were found in joint evaluation and forced choice, but not in separate evaluation. Joint evaluation increased both the weak and, to a lesser extent, the strong existence effect. Forced choice slightly increased the strong existence effect.

---

[14]The experimental design allowed me to also test the singularity effect (see footnote 1) in separate evaluation in both IVE-studies. Contradicting the results reported by Kogut and Ritov (2005b, Study 2), attractiveness ratings were higher for participants reading about 3 identified patients than for those reading about 1 identified patient, both when the patients were children, t(392) = 4.16, p <.001 (IVE1) and when the patients were adults t(470) = 5.81, p < .001 (IVE2). Resource allocations were slightly higher for those reading about 3 identified adults than for those reading about 1 identified adult in IVE2, t(470) = 2.69, p = .007, but not significantly so when the patients were children, t(392) = 0.86, p = .388 (IVE1). Despite not being a direct replication of previous singularity studies, the obtained results should cast doubt over the robustness of the singularity effect (see also Thomas-Walters & Raihani, 2017 and Wang et al., 2015).

TABLE 7: Results for the existence effect.

	Project A	Project B	Equal	Test	Percentage
Weak effect	**A(X)**	**B(X)**			
	6 now	**6 in one year**			
SE rating	72.79 (21.09)	69.95 (23.89)		t[447] = 1.34, p = .182	53.55%
SE allocation	45.39 (24.22)	44.02 (24.64)		t[447] = 0.59, p = .553	51.58%
JE rating	81.46 (19.17)	57.72 (22.99)		t[70] = 9.71, p < .001	87.54%
JE allocation	77.45 (13.03)	22.55 (13.03)		t[70] = 17.76, p < .001	98.24%
JE rating preference	66	3	2	z = 7.48, p < .001	94.37%
JE allocation preference	69	0	2	z = 8.19, p < .001	98.59%
Forced choice preference	67	1		z = 8.00, p < .001	98.53%
Strong effect	**A(X)**	**B(X-2)**			
	4 now	**6 in one year**			
SE rating	68.96 (24.69)	69.95 (23.89)		t[448] = −0.43, p = .668	48.85%
SE allocation	43.97 (23.97)	44.02 (24.64)		t[448] = −0.02, p = .983	49.94%
JE rating	80.08 (17.88)	70.59 (19.53)		t[70] = 4.21, p < .001	69.13%
JE allocation	59.99 (21.22)	40.01 (21.22)		t[70] = 3.97, p < .001	68.11%
JE rating preference	43	25	3	z = 2.14, p = .033	62.68%
JE allocation preference	39	18	14	z = 2.49, p = .013	64.79%
Forced choice preference	55	14		z = 4.94, p < .001	79.71%

3.4.1 Weak existence effect (6 patients now vs. 6 patients one year later)

Separate evaluation. Participants reading about a project helping six existing patients gave similar attractiveness ratings (M = 72.79 vs. 69.95), and earmarked similar amounts of resources (M = 45.39 vs. 44.02), compared to those reading about a project helping equally many patients one year later.

Joint evaluation. Yet, when evaluated side by side, the project helping six existing patients was rated as much more attractive than the project helping six patients one year later (M = 81.46 vs. 57.72), and also allocated more resources (M = 77.45 vs. 22.55).

The mean difference (Project A [6 existing] minus Project B [6 future]) in ratings was more than 20 points higher in joint evaluation (SE = 2.84, JE = 23.74). Likewise, the mean difference in allocations was more than 50 points higher in joint evaluation (SE = 1.37, JE = 54.90). This indicates that joint evaluation increases the weak existence effect.

Forced choice.   98.53% chose to help 6 patients now rather than 6 patients in one year. Preferences expressed with forced choice did not differ from preferences inferred from joint evaluation ratings ($\chi2$ = 1.74, p = .188) or allocations ($\chi2$ < 0.01, p = .975).

3.4.2   Strong existence effect (4 patients now vs. 6 patients one year later)

Separate evaluation.   Participants reading about a project helping four existing patients and participants reading about a project helping six future patients gave similar attractiveness ratings (M = 68.96 vs. 69.95), and earmarked similar amounts of resources (M = 43.97 vs. 44.02).

Joint evaluation.   Still, when evaluated side by side, the project helping four existing patients was rated as more attractive than the project helping six patients one year later (M = 80.08 vs. 70.59), and also allocated more resources (M= 59.99 vs. 40.01).

The mean difference (Project A [4 existing] minus Project B [6 future]) was more than 10 points higher in joint evaluation in ratings (SE = −0.99, JE = 9.49), and more than 20 points higher in allocations (SE = −0.05, JE = 19.98). This means that joint evaluation increases the strong existence effect as well.

Forced choice.   79.71% chose to help 4 patients now rather than 6 patients in one year. Participants were slightly more likely to express preferences in line with the strong existence effect in forced choice, than in attractiveness ratings ($\chi2$ = 4.82, p = .028) or resource allocations ($\chi2$ = 3.88, p = .049).

3.5   Age effect

The weak age effect was not found in separate evaluation, but it was found in joint evaluation and to an even greater extent in forced choice. The strong age effect was generally absent, but preferences expressed with forced choices were slightly more in favor of helping 4 children over 6 adults, than preferences inferred from attractiveness ratings.

3.5.1   Weak age effect (6 children vs. 6 adults)

Separate evaluation.   Participants reading about a project helping children and partici-pants reading about a project helping equally many adults gave similar attractiveness ratings (M = 68.97 vs. 67.45) and earmarked similar amounts of resources (M = 41.15 vs. 40.74).

Joint evaluation.   Still, when evaluated side by side, the project helping children was rated as slightly more attractive than the project helping equally many adults (M = 75.48 vs. 72.13), and also allocated more resources (M = 60.15 vs. 39.85).

TABLE 8: Results for the age effect.

	Project A	Project B	Equal	Test	Percentage
Weak effect	**A(X)**	**B(X)**			
	6 children	**6 adults**			
SE rating	68.97 (26.32)	67.45 (24.97)		t[433] = 0.62, p = .538	51.67%
SE allocation	41.15 (24.43)	40.74 (23.18)		t[433] = 0.18, p = .856	50.49%
JE rating	75.48 (21.83)	72.13 (23.65)		t[72] = 2.18, p = .032	60.06%
JE allocation	60.15 (12.98)	39.85 (12.98)		t[72] = 6.68, p < .001	78.29%
JE rating preference	33	16	24	z = 1.99, p = .047	61.64%
JE allocation preference	46	3	24	z = 5.03, p < .001	79.45%
Forced choice preference	63	8		z = 6.53. p < .001	88.73%
Strong effect	**A(X-2)**	**B(X)**			
	4 children	**6 adults**			
SE rating	65.38 (27.57)	67.45 (24.97)		t[432] = −0.82, p = .414	47.78%
SE allocation	39.99 (24.56)	40.74 (23.18)		t[432] = −0.33, p = .744	49.11%
JE rating	73.68 (21.11)	76.24 (20.35)		t[66] = −1.65, p = .103	42.01%
JE allocation	52.10 (18.00)	47.90 (18.00)		t[66] = 0.96, p = .342	54.64%
JE rating preference	19	35	13	z = −1.96, p = .051	38.06%
JE allocation preference	27	24	16	z = 0.37, p = .714	52.24%
Forced choice preference	41	27		z = 1.70, p = .090	60.29%

The mean difference (Project A [6 children] minus Project B [6 adults]) in ratings was about the same (SE = 1.52, JE = 3.35), but the mean difference in allocations was almost 20 points higher in joint evaluation (SE = 0.41, JE = 20.30).

Forced choice.   88.73% chose the project helping six children over the project helping six adults. Participants were slightly more likely to express preferences in line with the weak age effect in forced choice, than in attractiveness ratings ($\chi 2$ = 14.09, p < .001), but not than in resource allocations ($\chi 2$ = 2.31, p = .129).

3.5.2   Strong age effect (4 children vs. 6 adults)

Separate evaluation.   Participants reading about a project treating four children and participants reading about a project treating six adults gave similar attractiveness ratings (M = 65.38 vs. 67.45) and earmarked similar amounts of resources (M = 39.99 vs. 40.74).

1137

Joint evaluation.   When evaluated side by side, the project helping four children was rated as slightly less attractive than the project treating six adults (M = 73.68 vs. 76.24).[15] Still, the two projects were allocated about equal amounts of resources (M = 52.10 vs. 47.90).

The mean difference (Project A [4 children] minus Project B [6 adults]) in ratings was about the same (SE = −2.07, JE = −2.56). The mean difference in allocations was slightly larger in joint evaluation (SE = −0.75, JE = 4.20).

Forced choice.   Nevertheless, 60.29% chose to help four children rather than six adults. Participants were slightly more likely to express preferences in line with the strong age effect in forced choice, than in attractiveness ratings ($\chi2$ = 6.73, p = .009), but not than in resource allocations ($\chi2$ = 0.89, p = .345).

3.6   Innocence effect

The weak innocence effect was not found in separate evaluation, but it was found in joint evaluation and to a greater extent in forced choice. The strong innocence effect was generally absent, but preferences expressed with forced choices were slightly more in favor of helping 4 innocent gymmers rather than 6 non-innocent smokers, than preferences inferred from attractiveness ratings.

3.6.1   Weak innocence effect (6 "gymmers" vs. 6 "smokers")

Separate evaluation.   Participants reading about a project treating innocent patients and participants reading about a project treating equally many non-innocent patients gave similar attractiveness ratings (M = 63.80 vs. 65.41), and earmarked similar amounts of resources (M = 39.53 vs. 43.59).

Joint evaluation.   When evaluated side by side, the project helping six innocent patients was rated as slightly more attractive than the project helping six non-innocent patients (M = 70.29 vs. 62.96), and also allocated more resources (M = 58.49 vs. 41.51).

The mean difference (Project A [6 innocent] minus Project B [6 non-innocent]) was larger in joint evaluation both in ratings (SE = −1.61, JE = 7.33) and in allocations (SE = −4.06, JE = 16.98). This indicates that joint evaluation increases the weak innocence effect.

Forced choice.   81.25% chose to implement the project treating six innocent gymmers rather than the project treating six non-innocent smokers. Participants were more likely to express preferences in line with the weak innocence effect in forced choice, than in attractiveness ratings ($\chi2$ = 9.23, p = .002) or in resource allocations ($\chi2$ = 4.50, p = .034).

---

[15]This effect was significant when using a non-parametric test (see the supplement).

TABLE 9: Results for the innocence effect.

	Project A	Project B	Equal	Test	Percentage
Weak effect	**A(X)**	**B(X)**			
	6 gymmers	**6 smokers**			
SE rating	63.80 (25.30)	65.41 (23.75)		t[431] = −0.69, p = .494	48.15%
SE allocation	39.53 (22.30)	43.59 (24.17)		t[431] = −1.82, p = .070	45.09%
JE rating	70.29 (22.58)	62.96 (26.16)		t[84] = 2.12, p = .037	59.10%
JE allocation	58.49 (19.05)	41.51 (19.05)		t[84] = 4.11, p < .001	67.21%
JE rating preference	46	30	9	z = 1.74, p = .083	59.41%
JE allocation preference	46	18	21	z = 3.04, p = .002	66.47%
Forced choice preference	65	15		z = 5.59, p < .001	81.25%
Strong effect	**A(X-2)**	**B(X)**			
	4 gymmers	**6 smokers**			
SE rating	64.75 (25.95)	65.41 (23.75)		t[429] = −0.28, p = .782	49.25%
SE allocation	44.06 (24.48)	43.59 (24.17)		t[429] = 0.20, p = .843	50.54%
JE rating	62.81 (22.01)	69.30 (18.42)		t[80] = −2.28, p = .025	40.00%
JE allocation	49.51 (16.73)	50.49 (16.73)		t[80] = −0.27, p = .791	48.83%
JE rating preference	28	46	7	z = −2.00, p = .046	38.89%
JE allocation preference	33	31	17	z = 0.22, p = .825	51.23%
Forced choice preference	53	37		z = 1.69, p = .092	58.89%

3.6.2   Strong innocence effect (4 "gymmers" vs. 6 "smokers")

Separate evaluation.   Participants reading about a project treating four innocent patients and participants reading about a project treating six non-innocent patients gave similar attractiveness ratings (M = 64.75 vs. 65.41), and earmarked similar amounts of resources (M = 44.06 vs. 43.59).

Joint evaluation.   When evaluated side by side, the project helping four innocent patients was rated as slightly less attractive than the project treating six non-innocent patients (M = 62.81 vs. 69.30). Still, the two projects were allocated about equal amounts of resources (M = 49.51 vs. 50.49).

   The mean difference (Project A [4 innocent] minus Project B [6 non-innocent]) in ratings was slightly lower (more negative) in joint than in separate evaluation (SE = −0.66, JE = −6.49). The mean difference in allocations was about the same (SE = 0.43, JE = −0.98).

Forced choice. Despite this, 58.89% chose to implement the project helping four gymmers rather than the project helping six smokers. Participants were slightly more likely to express preferences in line with the strong innocence effect in forced choice, than in attractiveness ratings ($\chi 2 = 6.87$, $p = .009$), but not than in resource allocations ($\chi 2 = 1.00$, $p = .317$).

3.7 Gender effect

The weak gender effect was not found in separate or joint evaluation, but it clearly appeared in forced choice. The strong gender effect was not found in any decision mode. On the contrary, participants expressed clear preferences for treating more males rather than fewer females in joint evaluation and in forced choice (i.e., a reversed strong gender effect).

TABLE 10: Results for the gender effect.

	Project A	Project B	Equal	Test	Percentage
Weak effect	**A(X)**	**B(X)**			
	6 women	**6 men**			
SE rating	70.16 (26.57)	70.72 (22.96)		t[442] = −0.24, p = .813	49.37%
SE allocation	41.71 (24.91)	42.86 (24.03)		t[442] = −0.49, p = .622	48.68%
JE rating	70.69 (22.68)	69.71 (23.32)		t[74] = 1.50, p = .138	56.86%
JE allocation	50.72 (4.01)	49.28 (4.01)		t[74] = 1.56, p = .124	57.12%
JE rating preference	22	19	34	z = 0.35, p = .729	52.00%
JE allocation preference	9	3	63	z = 0.69, p < .488	54.00%
Forced choice preference	62	18		z = 4.92, p < .001	77.50%
Strong effect	**A(X-2)**	**B(X)**			
	4 women	**6 men**			
SE rating	66.81 (27.00)	70.72 (22.96)		t[439] = −1.63, p = .103	45.61%
SE allocation	40.94 (24.77)	42.86 (24.03)		t[439] = −0.82, p = .410	47.78%
JE rating	65.59 (21.58)	73.48 (19.71)		t[73] = −5.67, p < .001	25.47%
JE allocation	43.65 (9.94)	56.35 (9.94)		t[73] = −5.47, p < .001	26.15%
JE rating preference	10	57	7	z = −5.46, p < .001	18.24%
JE allocation preference	9	45	20	z = −4.18, p < .001	25.68%
Forced choice preference	23	52		z = −3.35, p < .001	30.67%

3.7.1   Weak gender effect (6 females vs. 6 males)

Separate evaluation.   Participants reading about a project helping females and participants reading about an otherwise identical project helping equally many males gave similar attractiveness ratings (M = 70.16 vs. 70.72), and earmarked similar amounts of resources (M = 41.71 vs. 42.86).

Joint evaluation.   When evaluated side by side, participants rated the project helping six females and the project helping six males as similarly attractive (M = 70.69 vs. 69.71), and allocated resources evenly between the two projects (M = 50.72 vs. 49.28).

The mean difference (Project A [6 females] minus Project B [6 males]) was similar and around zero in both separate and joint evaluation for ratings (SE = −0.56, JE = 0.98) and for allocations (SE = −1.15, JE = 1.44).

Forced choice.   Still, 77.50% chose the project helping six females over the project helping six males.[16] Participants were more likely to express preferences in line with the weak gender effect in forced choice, than in attractiveness ratings ($\chi2$ = 11.09, p < .001), or in resource allocations ($\chi2$ = 9.45, p = .002).

3.7.2   Strong gender effect (4 females vs. 6 males)

Separate evaluation.   Participants reading about a project helping four females and participants reading about a project helping six males gave similar attractiveness ratings (M = 66.81 vs. 70.72), and earmarked similar amounts of resources (M = 40.94 vs. 42.86).

Joint evaluation.   When evaluated side by side, the project helping six males was rated as more attractive than the project helping four females (M = 65.59 vs 73.48), and also allocated more resources (M = 43.65 vs. 56.35).

The mean difference (Project A [4 females] minus Project B [6 males]) in attractiveness ratings was slightly lower (more negative) in joint evaluation (SE = −3.91, JE = −7.89). Likewise the mean difference in allocations was almost 11 points lower in joint evaluation (SE = −1.92, JE = −12.70). This indicates that joint evaluation reverses the strong gender effect.

Forced choice.   30.67% chose to help four females rather than six males.[17] Preferences expressed with forced choice did not differ much from preferences inferred from joint evaluation ratings ($\chi2$ = 3.32, p = .068) or from allocations ($\chi2$ = 0.46, p = .498).

---

[16]This tendency was significantly different from 50–50 among female participants (86.8%, z = 4.54, p < .001) as well as among male participants (68.3%, z = 2.34, p = .019).

[17]The percentage choosing four females was significantly different from 50% among male participants (23.9%, z = 2.04, p = .041) but not among female participants (37.8%, z = 1.48, p = .139).

4   General discussion

The weak and strong forms of seven helping effects were systematically tested in three decision modes (separate evaluation, joint evaluation and forced choice) using a unified experimental paradigm and with over 9000 participants. I think there are at least three lessons to learn from this research.

The first lesson is that many helping effects are notably difficult to find in separate evaluation. When evaluated one at the time, projects helping children, innocent patients, existing patients, and fellow citizens were rated as no more attractive and allocated no more resources than identical projects helping equally many adults, "non-innocent" patients (smokers), future patients, and foreigners. This is noteworthy considering that most of these effects clearly emerged in joint evaluation.

The IGE-family and the IVE did emerge to some extent also in separate evaluation, but the PDE was the effect that clearly stood out. In both studies, people rated a high rescue proportion project as more attractive and allocated it more resources than people who read about a low rescue proportion project. This suggests that numerical attributes can influence moral preferences more than categorical attributes in separate evaluation, but only when the numbers are easily evaluable (e.g., expressed as proportions rather than absolute numbers; Bartels, 2006; Hsee & Zhang, 2010). In the terms used by Li and Hsee (2019), these results show that the seven included attributes differ in their relative evaluability. To exemplify, age and innocence of the beneficiary as well as existence are relatively difficult to evaluate in isolation, whereas rescue proportion, and to lesser extent identifiability and family-belonging, are relatively easy to evaluate.

The second lesson is that the weak vs. the strong forms of helping effects elicited similar response patterns in separate evaluation, but quite different patterns in joint evaluation, for some of the effects. This confirms the assumption that the efficiency attribute (number of people possible to treat) matters for people, but is difficult to evaluate in isolation (similar to the number of entries in a dictionary; Hsee, 1996). A novel finding was that the relative importance of the efficiency attribute (compared to the contrasting attribute) differed much across helping effects. The weak IVE was, e.g., found in joint evaluation (people preferred 3 identified over 3 non-identified) but there were large reversed effects when testing its strong form (3 non-identified much preferred over 1 identified). This result suggests that the number-of-victims attribute is more justifiable than the identifiability attribute.

Likewise, the IGE, age and innocence effects appeared in their weak form in joint evaluation, but were absent or reversed in their strong form, whereas the PDE and especially the existence effect were found also in their strong forms. The gender effect was instead absent in its weak form (6 females equally preferred as 6 males) but reversed in its strong form (6 males preferred over 4 females). Together, these results suggest that people value both the number of individuals possible to save, but also other attributes. The identifiability, nationality and gender attributes have relatively low justifiability, and are thus easily outweighed by the number-of-patients attribute (which was held constant across studies),

1142

whereas it is somehow easier for people to justify helping fewer existing over more future patients. Further elucidating why some attributes fare better and other fare worse when pitted against an efficiency attribute should be a prioritized research area in the future.[18]

The third lesson is that preferences inferred from attractiveness ratings and resource allocations in joint evaluation do not always correspond with preferences obtained from forced choices, despite that these decision modes share the joint evaluation feature (Erlandsson et al., 2020; Slovic, 1975). The clearest evidence of this is found in the weak gender effect where participants expressed no preference between saving 6 male and 6 female patients when expressed with ratings or resource-allocations, but a robust preference for helping females when they were forced to choose. One explanation for this is that participants are first and foremost motivated to express justifiable moral preferences (Capraro & Rand, 2018; Choshen-Hillel et al., 2015). The most easily justifiable preference is to claim that males and females are equally valuable, so most people do so in rating and allocation tasks. In the choice task it was impossible to express indifference, but rather than then choosing randomly (which a truly indifferent decision-maker would do), it is possible that people then go for the second most justifiable preference which is to value females higher than males.

Additional support for that the forced choice decision mode influences preferences was found in other helping effects. Compared to preferences inferred from attractiveness ratings, forced choice made people more in favor of helping ingroup rather than outgroup members (weak and strong IGE), helping children rather than adults (weak and strong age effect), helping fewer existing patients rather than more future patients (strong existence effect) and helping fewer innocent patients rather than more non-innocent patients (strong innocence effect). The opposite pattern was found for the strong PDE where forced choice made people slightly more in favor of a low rescue proportion project helping more patients (e.g., 6 of 100), rather than a high proportion project helping fewer patients (e.g., 4 of 5).

The prominence effect (Slovic, 1975; Tversky et al., 1988) argues that the relatively more prominent (important) attribute influences choices more than it influence other types of joint evaluation preference expressions (e.g., attractiveness-ratings, allocations, or contingent valuation). In this light, the results reported here suggest that innocent patients, existing patients, children and especially one's ingroup are more prominent attributes than the number of victims possible to save, whereas rescue proportion is less prominent.

4.1 Limitations

I am not suggesting that the unified paradigm used by me (hypothetical medical helping projects presented in tabular form) is superior to other unified paradigms that could test

---

[18]It is again worth emphasizing that the aim of this study was to compare weak helping effects (where the number of patients possible to help is equal) against "minimal" strong helping effects (where the number of patients possible to help is similar but unequal). I therefore opted for the "A(X) vs B(X)" and "A(X) vs B(X-2)" paradigm in all 10 studies and purposely kept X small (6 or 3 patents). The results would most likely turn out different if X was larger (e.g., 1000 outgroup vs. 998 ingroup patients) and if the difference in efficiency was larger (e.g., 1000 outgroup vs. 2 ingroup patients).

the same effects. Helping effects are unavoidably context-dependent so it is possible that some effects are more or less "easy to find" using different paradigms. I welcome not only direct replications, but also conceptual replications of these studies in order to determine how generalizable the results are.[19]

Three related limitations worth mentioning are: (1) That the difference in the efficiency attribute was small when testing the strong helping effects (6 vs. 4 patients or 3 vs. 1 patient). (2) That this study suffers from the W.E.I.R.D-problem, as all participants were English-speaking and recruited from MTurk or Prolific, and thus not representative of the global population (Henrich et al., 2010). (3) That all helping effects were tested using non-behavioral outcome variables. Possible ways to further improve helping-effect research include investigating where the efficiency related tipping points are for different effects (e.g., how many future patients must be treated in order to surpass treating one existing patient; Dolan & Tsuchiya, 2011; Erlandsson et al., 2020), investigating cultural, demographic and personality-based differences in both weak and strong helping effects (e.g., Deshpande & Spears, 2016; Fiedler, et al., 2018; Wang et al., 2015), and investigating whether different effects are differently affected when moving from hypothetical helping decisions to real (cost-incurring) helping decisions (Ferguson et al., 2019).

It is worth noting that results from attractiveness ratings did not always correspond with results from resource allocations. Differences in joint evaluation preferences were always the largest between ratings and forced choices (with allocation located somewhere between), and this pattern of results suggests that resource allocations represent something that is located between attractiveness-ratings and forced choices in terms of decision modes. Expressed differently, allocations are more "choice-like" than ratings, but more "rating-like" than choices.

Lastly, a potential experimental confound is that the joint evaluation and forced choice modes differ not only in possibility to express indifference (possible vs. impossible) but also in type of elicitation task (rating/allocation vs. choice). It is worth pointing out that one can manipulate the possibility to express indifference both in rating tasks (e.g., by making it possible or impossible to give equal ratings) allocations (by having people allocate even or uneven amounts) and in choice-tasks (by giving or not giving participants the option to pass on the choice to someone else or to opt out from choosing all together).

4.2 Conclusion

The main insight from this paper is that helping effects can be tested in different ways, that different effects are differently affected when moving from one type of test to another, and

---

[19]There is also room to make the relatively unified paradigm even more unified and refined. In hindsight, it would, e.g., have been preferable to consistently use the same number of patients in all ten studies, rather than using 4 vs. 6 in some studies and 1 vs 3 in others, and to test the IVE using the same layout as in the other studies (or vice versa). In would also have been preferable to use 50% (rather than 20%) as the anchor in separate evaluation allocations, to make the separate and joint allocation tasks more similar.

that these different response patterns can be understood as help-situation attributes differing in their evaluability, justifiability, and prominence. Some helping effects are present in joint evaluation but absent in separate evaluation whereas other effects give rise to the opposite pattern, and yet other effects are found only when people cannot express indifference. My hope is that this article can inspire researchers to routinely investigate also other helping effect in their weak and strong presentational form and in multiple decision modes, as this will provide us with a more nuanced and multi-faceted perspective of the psychology of prosocial decision making.

References

Ahn, H.-K., Kim, H. J., & Aggarwal, P. (2014). Helping fellow beings: Anthropomorphized social causes and the role of anticipatory guilt. *Psychological Science, 25*(1), 224–229. https://doi.org/10.1177/0956797613496823.

Back, S., & Lips, H. M. (1998). Child sexual abuse: Victim age, victim gender, and observer gender as factors contributing to attributions of responsibility. *Child Abuse & Neglect, 22*(12), 1239–1252. https://doi.org/10.1016/S0145-2134(98)00098-2.

Baron, J. (1997). Confusion of relative and absolute risk in valuation. *Journal of Risk and Uncertainty, 14*(3), 301–309. https://doi.org/10.1023/A:1007796310463.

Baron, J. (2012). Parochialism as a result of cognitive biases. In R. Goodman, D. Jinks, & A. K. Woods (Eds.), *Understanding social action, promoting human rights*, pp. 203–243. Oxford: Oxford University Press.

Baron, J., Ritov, I., & Greene, J. D. (2013). The duty to support nationalistic policies. *Journal of Behavioral Decision Making, 26*(2), 128–138. https://doi.org/10.1002/bdm.768.

Baron, J., & Szymanska, E. (2011). Heuristics and biases in charity. In D. M. Oppenheimer & C. Y. Olivola (Eds.), *The Science of Giving: Experimental Approaches to the Study of Charity.* (pp. 215-235). New York, NY, US: Psychology Press.

Bartels, D. M. (2006). Proportion dominance: The generality and variability of favoring relative savings over absolute savings. *Organizational Behavior and Human Decision Processes, 100*(1), 76–95. https://doi.org/10.1016/j.obhdp.2005.10.004.

Bartels, D. M., & Burnett, R. C. (2011). A group construal account of drop-in-the-bucket thinking in policy preference and moral judgment. *Journal of Experimental Social Psychology, 47*(1), 50–57. https://doi.org/10.1016/j.jesp.2010.08.003.

Bazerman, M. H., Gino, F., Shu, L. L., & Tsay, C.-J. (2011). Joint evaluation as a real-world tool for managing emotional assessments of morality. *Emotion review, 3*(3), 290–292. https://doi.org/10.1177/1754073911402370.

Bazerman, M. H., Moore, D. A., Tenbrunsel, A. E., Wade-Benzoni, K. A., & Blount, S. (1999). Explaining how preferences change across joint versus separate evaluation.

Journal of Economic Behavior & Organization, 39(1), 41–58. https://doi.org/10.1016/S0167-2681(99)00025-6.

Bischoff, C., & Hansen, J. (2016). Influencing support of charitable objectives in the near and distant future: Delay discounting and the moderating influence of construal level. *Social Influence, 11*(4), 217-229. https://doi.org/10.1080/15534510.2016.1232204.

Bohnet, I., Van Geen, A., & Bazerman, M. (2016). When performance trumps gender bias: Joint vs. separate evaluation. *Management Science*, *62*(5), 1225–1234. https://doi.org/10.1287/mnsc.2015.2186.

Bradley, A., Lawrence, C., & Ferguson, E. (2019). When the relatively poor prosper: the underdog effect on charitable donations. *Nonprofit and Voluntary Sector Quarterly, 48*(1), 108–127. https://doi.org/10.1177/0899764018794305.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues, 55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126.

Broome, J. (1984). Selecting people randomly. *Ethics, 95*(1), 38–55. https://doi.org/10.1086/292596.

Burnstein, E., Crandall, C., & Kitayama, S. (1994). Some neo-Darwinian decision rules for altruism: Weighing cues for inclusive fitness as a function of the biological importance of the decision. *Journal of Personality and Social Psychology, 67*(5), 773–789. https://doi.org/10.1037/0022-3514.67.5.773.

Butts, M. M., Lunt, D. C., Freling, T. L., & Gabriel, A. S. (2019). Helping one or helping many? A theoretical integration and meta-analytic review of the compassion fade literature. *Organizational Behavior and Human Decision Processes, 151*, 16–33. https://doi.org/10.1016/j.obhdp.2018.12.006.

Capraro, V., & Rand, D. G. (2018). Do the right thing: Experimental evidence that preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality. *Judgment and Decision Making, 13*(1), 99–111. http://journal.sjdm.org/17/171107/jdm171107.html.

Carroll, L. S., White, M. P., & Pahl, S. (2011). The impact of excess choice on deferment of decisions to volunteer. *Judgment and Decision Making, 6*(7), 629–637. http://journal.sjdm.org/11/11418/jdm11418.html.

Caviola, L., Faulmüller, N., Everett, J., Savulescu, J., & Kahane, G. (2014). The evaluability bias in charitable giving: Saving administration costs or saving lives? *Judgment and Decision Making, 9*(4), 303–315. http://www.sjdm.org/journal/14/14402a/jdm14402a.html.

Caviola, L., Schubert, S., & Nemirow, J. (2020). The many obstacles to effective giving. *Judgment and Decision Making, 15*(2), 159–172. http://journal.sjdm.org/19/190810/jdm190810.html.

Chapman, G. B., & Elstein, A. S. (1995). Valuing the future: Temporal discounting of health and money. *Medical Decision Making, 15*(4), 373–386. https://doi.org/10.1177/0272989X9501500408.

Choshen-Hillel, S., Shaw, A., & Caruso, E. M. (2015). Waste management: How reducing partiality can promote efficient resource allocation. *Journal of Personality and Social Psychology, 109*(2), 210–231. https://psycnet.apa.org/doi/10.1037/pspa0000028

Cropper, M. L., Aydede, S. K., & Portney, P. R. (1994). Preferences for life saving programs: how the public discounts time and age. *Journal of Risk and Uncertainty, 8*(3), 243–265. https://doi.org/10.1007/BF01064044.

Curry, T. R., Lee, G., & Rodriguez, S. F. (2004). Does victim gender increase sentence severity? Further explorations of gender dynamics and sentencing outcomes. *Crime & Delinquency, 50*(3), 319–343. https://doi.org/10.1177/0011128703256265

De Dreu, C. K., Greer, L. L., Van Kleef, G. A., Shalvi, S., & Handgraaf, M. J. (2011). Oxytocin promotes human ethnocentrism. *Proceedings of the National Academy of Sciences, 108*(4), 1262–1266. https://doi.org/10.1073/pnas.1015316108.

Deshpande, A., & Spears, D. (2016). Who is the identifiable victim? Caste and charitable giving in modern India. *Economic Development and Cultural Change, 64*(2), 299–321. https://doi.org/10.1086/684000.

Dickert, S., Kleber, J., Västfjäll, D., & Slovic, P. (2016). Mental imagery, impact, and affect: A mediation model for charitable giving. *PloS one, 11*(2). https://doi.org/10.1371/journal.pone.0148274.

Dickert, S., & Slovic, P. (2009). Attentional mechanisms in the generation of sympathy. *Judgment and Decision Making, 4*(4), 297–306. http://www.sjdm.org/journal/9417/jdm9417.html.

Dickert, S., Västfjäll, D., Kleber, J., & Slovic, P. (2012). Valuations of human lives: normative expectations and psychological mechanisms of (ir)rationality. *Synthese*, 1–11. https://doi.org/10.1007/s11229-012-0137-4.

Dolan, P., & Tsuchiya, A. (2011). Determining the parameters in a social welfare function using stated preference data: An application to health. *Applied Economics, 43*(18), 2241–2250. https://doi.org/10.1080/00036840903166244

Duclos, R., & Barasch, A. (2014). Prosocial behavior in intergroup relations: How donor self-construal and recipient group-membership shape generosity. *Journal of Consumer Research, 41*(1), 93–108. https://doi.org/10.1086/674976.

Dufwenberg, M., & Muren, A. (2006). Generosity, anonymity, gender. *Journal of Economic Behavior & Organization, 61*(1), 42–49. https://doi.org/10.1016/j.jebo.2004.11.007.

Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin, 100*(3), 283–308. https://psycnet.apa.org/doi/10.1037/0033-2909.100.3.283.

Ein-Gar, D., & Levontin, L. (2013). Giving from a distance: Putting the charitable organization at the center of the donation appeal. *Journal of Consumer Psychology, 23*(2), 197–211. https://doi.org/10.1016/j.jcps.2012.09.002.

Erlandsson, A., Björklund, F., & Bäckström, M. (2014). Perceived utility (not sympathy) mediates the proportion dominance effect in helping decisions. *Journal of Behavioral*

Decision Making, 27(1), 37–47. https://doi.org/10.1002/bdm.1789.

Erlandsson, A., Björklund, F., & Bäckström, M. (2015). Emotional reactions, perceived impact and perceived responsibility mediate the identifiable victim effect, proportion dominance effect and in-group effect respectively. *Organizational Behavior and Human Decision Processes, 127*(0), 1–14. https://doi.org/10.1016/j.obhdp.2014.11.003.

Erlandsson, A., Björklund, F., & Bäckström, M. (2017). Choice-justifications after allocating resources in helping dilemmas. *Judgment and Decision Making, 12*(1), 60–80. http://journal.sjdm.org/15/15410/jdm15410.html.

Erlandsson, A., Lindkvist, A., Lundqvist, K., Andersson, P. A., Dickert, S., Slovic, P., & Västfjäll, D. (2020). Moral preferences in helping dilemmas expressed by matching and forced choice. *Judgment and Decision Making, 15*(4), 452–475. http://journal.sjdm.org/20/200428/jdm200428.html.

Erlandsson, A., Nilsson, A., & Västfjäll, D. (2018). Attitudes and donation behavior when reading positive and negative charity appeals. *Journal of Nonprofit & Public Sector Marketing*, 1–31. https://doi.org/10.1080/10495142.2018.1452828.

Everett, J. A. C., Faber, N. S., & Crockett, M. (2015). Preferences and beliefs in ingroup favoritism. *Frontiers in Behavioral Neuroscience, 9*(15). https://doi.org/10.3389/fnbeh.2015.00015.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146.

Fehse, K., Silveira, S., Elvers, K., & Blautzik, J. (2015). Compassion, guilt and innocence: An fMRI study of responses to victims who are responsible for their fate. *Social Neuroscience, 10*(3), 243–252. https://doi.org/10.1080/17470919.2014.980587

Feiler, D. C., Tost, L. P., & Grant, A. M. (2012). Mixed reasons, missed givings: The costs of blending egoistic and altruistic reasons in donation requests. *Journal of Experimental Social Psychology, 48*(6), 1322—1328. https://doi.org/10.1016/j.jesp.2012.05.014.

Ferguson, E., Zhao, K., O'Carroll, R. E., & Smillie, L. D. (2019). Costless and costly prosociality: correspondence among personality traits, economic preferences, and real-world prosociality. *Social Psychological and Personality Science, 10*(4), 461–471. https://doi.org/10.1177/1948550618765071.

Fetherstonhaugh, D., Slovic, P., Johnson, S., & Friedrich, J. (1997). Insensitivity to the value of human life: A study of psychophysical numbing. *Journal of Risk and Uncertainty, 14*(3), 283–300. https://doi.org/10.1023/a:1007744326393.

Fiedler, S., Hellmann, D. M., Dorrough, A. R., & Glöckner, A. (2018). Cross-national ingroup favoritism in prosocial behavior: Evidence from Latin and North America. *Judgment & Decision Making, 13*(1), 42–60. http://journal.sjdm.org/17/17818a/jdm17818a.html.

Fischer, G. W., & Hawkins, S. A. (1993). Strategy compatibility, scale compatibility, and the prominence effect. *Journal of Experimental Psychology: Human Perception and*

Performance, *19*(3), 580–597. https://doi.org/https://doi.org/10.1037/0096-1523.19.3. 580.

Fisher, R. J., & Ma, Y. (2014). The price of being beautiful: Negative effects of attractiveness on empathy for children in need. *Journal of Consumer Research, 41*(2), 436–450. https://doi.org/10.1086/676967.

Fong, C. M. (2007). Evidence from an experiment on charity to welfare recipients: Reciprocity, altruism and the empathic responsiveness hypothesis. *The Economic Journal, 117*(522), 1008–1024. https://doi.org/10.1111/j.1468-0297.2007.02076.x.

Friedrich, J., Barnes, P., Chapin, K., Dawson, I., Garst, V., & Kerr, D. (1999). Psychophysical numbing: When lives are valued less as the lives at risk increase. *Journal of Consumer Psychology, 8*(3), 277–299. https://doi.org/10.1207/s15327663jcp0803_05.

Friedrich, J., & McGuire, A. (2010). Individual differences in reasoning style as a moderator of the identifiable victim effect. *Social Influence, 5*(3), 182–201. https://doi.org/10.1080/15534511003707352.

Garinther, A., Arrow, H., & Razavi, P. (2021). Victim number effects in charitable giving: Joint evaluations promote egalitarian decisions. *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/0146167220982734.

Genevsky, A., Västfjäll, D., Slovic, P., & Knutson, B. (2013). Neural underpinnings of the identifiable victim effect: Affect shifts preferences for giving. *The Journal of Neuroscience, 33*(43), 17188–17196. https://doi.org/10.1523/jneurosci.2348-13.2013.

Goodwin, G. P., & Landy, J. F. (2014). Valuing different human lives. *Journal of Experimental Psychology: General, 143*(2), 778–803. https://doi.org/10.1037/a0032796.

Hart, P. S., Lane, D., & Chinn, S. (2018). The elusive power of the individual victim: Failure to find a difference in the effectiveness of charitable appeals focused on one compared to many victims. *PloS one, 13*(7). https://doi.org/10.1371/journal.pone.0199535.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29. https://doi.org/10.1038/466029a.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes, 67*(3), 247–257. https://doi.org/10.1006/obhd.1996.0077.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin, 125*(5), 576–590. https://doi.org/10.1037/0033-2909. 125.5.576.

Hsee, C. K., & Zhang, J. (2004). Distinction Bias: Misprediction and Mischoice Due to Joint Evaluation. *Journal of Personality and Social Psychology, 86*(5), 680–695. https:// doi.org/10.1037/0022-3514.86.5.680.

Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science, 5*(4), 343–355. https://doi.org/10.1177/1745691610374586.

Hsee, C. K., Zhang, J., Wang, L., & Zhang, S. (2013). Magnitude, time, and risk differ similarly between joint and single evaluations. *Journal of Consumer Research*, *40*(1), 172–184. https://doi.org/10.1086/669484.

Huber, M., Van Boven, L., McGraw, A. P., & Johnson-Graham, L. (2011). Whom to help? Immediacy bias in judgments and decisions about humanitarian aid. *Organizational Behavior and Human Decision Processes, 115*(2), 283–293. https://doi.org/10.1016/j.obhdp.2011.03.003.

James, T. K., & Zagefka, H. (2017). The effects of group memberships of victims and perpetrators in humanly caused disasters on charitable donations to victims. *Journal of Applied Social Psychology, 47*(8), 446–458. https://doi.org/10.1111/jasp.12452.

Jenni, K., & Loewenstein, G. (1997). Explaining the identifiable victim effect. *Journal of Risk and Uncertainty, 14*(3), 235–257. https://doi.org/10.1023/a:1007740225484.

Kahneman, D., Ritov, I., Jacowitz, K. E., & Grant, P. (1993). Stated willingness to pay for public goods: A psychological perspective. *Psychological Science*, *4*(5), 310–315. https://doi.org/10.1111/j.1467-9280.1993.tb00570.x.

Keren, G., & Teigen, K. H. (2010). Decisions by coin toss: Inappropriate but fair. *Judgment and Decision Making, 5*(2), 83–101. http://www.sjdm.org/~baron/journal/10/10203/jdm10203.html.

Kleber, J., Dickert, S., Peters, E., & Florack, A. (2013). Same numbers, different meanings: How numeracy influences the importance of numbers for pro-social behavior. *Journal of Experimental Social Psychology, 49*(4), 699–705. http://doi.org/10.1016/j.jesp.2013.02.009

Kogut, T. (2011). Someone to blame: When identifying a victim decreases helping. *Journal of Experimental Social Psychology, 47*(4), 748–755. https://doi.org/10.1016/j.jesp.2011.02.011.

Kogut, T., & Kogut, E. (2013). Exploring the relationship between adult attachment style and the identifiable victim effect in helping behavior. *Journal of Experimental Social Psychology, 49*(4), 651–660. https://doi.org/10.1016/j.jesp.2013.02.011.

Kogut, T., & Ritov, I. (2005a). The "Identified Victim" effect: An identified group, or just a single individual? *Journal of Behavioral Decision Making, 18*(3), 157–167. https://doi.org/10.1002/bdm.492.

Kogut, T., & Ritov, I. (2005b). The singularity effect of identified victims in separate and joint evaluations. *Organizational Behavior and Human Decision Processes, 97*(2), 106–116. https://doi.org/10.1016/j.obhdp.2005.02.003

Kogut, T., & Ritov, I. (2007). "One of us": Outstanding willingness to help save a single identified compatriot. *Organizational Behavior and Human Decision Processes, 104*(2), 150–157. https://doi.org/10.1016/j.obhdp.2007.04.006.

Kogut, T., & Ritov, I. (2015). Target dependent ethics: discrepancies between ethical decisions toward specific and general targets. *Current Opinion in Psychology, 6*, 145–149. https://doi.org/10.1016/j.copsyc.2015.08.005.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863.

Lee, S., & Feeley, T. H. (2016). The identifiable victim effect: a meta-analytic review. *Social Influence, 11*(3), 199–215. https://doi.org/10.1080/15534510.2016.1216891.

Lee, S., Winterich, K. P., & Ross, W. T. (2014). I'm moral, but I won't help you: The distinct roles of empathy and justice in donations. *Journal of Consumer Research, 41*(3), 678–696. https://doi.org/10.1086/677226.

Lesner, T. H., & Rasmussen, O. D. (2014). The identifiable victim effect in charitable giving: evidence from a natural field experiment. *Applied Economics, 46*(36), 4409–4430. https://doi.org/10.1080/00036846.2014.962226.

Levine, M., & Thompson, K. (2004). Identity, place, and bystander intervention: Social categories and helping after natural disasters. *The Journal of Social Psychology, 144*(3), 229–245. https://doi.org/10.3200/socp.144.3.229-245.

Li, M., & Chapman, G. B. (2009). "100% of anything looks good": The appeal of one hundred percent. *Psychonomic Bulletin & Review, 16*(1), 156–162. https://doi.org/10.3758/PBR.16.1.156.

Li, M., Vietri, J., Galvani, A. P., & Chapman, G. B. (2010). How do people value life? *Psychological Science, 21*(2), 163–167. https://doi.org/10.1177/0956797609357707.

Li, X., & Hsee, C. K. (2019). Beyond preference reversal: Distinguishing justifiability from evaluability in joint versus single evaluations. *Organizational Behavior and Human Decision Processes, 153*, 63–74. https://doi.org/10.1016/j.obhdp.2019.04.007.

Mata, A. (2016). Proportion dominance in valuing lives: The role of deliberative thinking. *Judgment and Decision Making, 11*(5), 441–448. http://journal.sjdm.org/15/15908/jdm15908.html.

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111*(2), 361–365. https://psycnet.apa.org/doi/10.1037/0033-2909.111.2.361.

Nowlis, S. M., & Simonson, I. (1997). Attribute–task compatibility as a determinant of consumer preference reversals. *Journal of Marketing Research*, *34*(2), 205–218. https://doi.org/https://doi.org/10.1177/002224379703400202.

O'Donoghue, T., & Rabin, M. (2015). Present bias: Lessons learned and to be learned. *American Economic Review*, *105*(5), 273–279. https://doi.org/10.1257/aer.p20151085.

Paharia, N., Kassam, K. S., Greene, J. D., & Bazerman, M. H. (2009). Dirty work, clean hands: The moral psychology of indirect agency. *Organizational Behavior and Human Decision Processes, 109*(2), 134–141. https://doi.org/10.1016/j.obhdp.2009.03.002.

Paolacci, G., & Yalcin, G. (2020). Fewer but poorer: Benevolent partiality in prosocial preferences. *Judgment and Decision Making, 15*(2), 173–181. http://sjdm.org/journal/19/191220a/jdm191220a.html.

Perrault, E. K., Silk, K. J., Sheff, S., Ahn, J., Hoffman, A., & Totzkay, D. (2015). Testing the identifiable victim effect with both animal and human victims in anti-littering messages. *Communication Research Reports, 32*(4), 294–303. https://doi.org/10.1080/08824096.2015.1089857.

Raihani, N. J., & Smith, S. (2015). Competitive helping in online giving. *Current Biology, 25*(9), 1183–1186. https://doi.org/10.1016/j.cub.2015.02.042.

Ritov, I., & Baron, J. (2011). Joint presentation reduces the effect of emotion on evaluation of public actions. *Cognition and Emotion, 25*(4), 657–675. https://doi.org/10.1080/02699931.2010.512512.

Sah, S., & Loewenstein, G. (2012). More affected = more neglected: Amplification of bias in advice to the unidentified and many. *Social Psychological and Personality Science, 3*(3), 365–372. https://doi.org/10.1177/1948550611422958.

Samuelson, P. A. (1937). A note on measurement of utility. *The Review of Economic Studies, 4*(2), 155–161. https://doi.org/10.2307/2967612

Schwartz-Shea, P., & Simmons, R. T. (1991). Egoism, parochialism, and universalism: Experimental evidence from the layered prisoners' dilemma. *Rationality and Society*, *3*(1), 106–132. https://doi.org/10.1177/1043463191003001007.

Seacat, J. D., Hirschman, R., & Mickelson, K. D. (2007). Attributions of HIV onset controllability, emotional reactions, and helping intentions: Implicit effects of victim sexual orientation. *Journal of Applied Social Psychology, 37*(7), 1442–1461. https://doi.org/10.1111/j.1559-1816.2007.00220.x

Sharps, D. L., & Schroeder, J. (2019). The preference for distributed helping. *Journal of Personality and Social Psychology, 117*(5), 954–977. https://doi.org/10.1037/pspi0000179.

Skedgel, C. D., Wailoo, A. J., & Akehurst, R. L. (2015). Choosing vs. allocating: Discrete choice experiments and constant-sum paired comparisons for the elicitation of societal preferences. *Health Expectations*, *18*(5), 1227–1240. https://doi.org/10.1111/hex.12098

Slovic, P. (1975). Choice between equally valued alternatives. *Journal of Experimental Psychology: Human Perception and Performance, 1*(3), 280–287. https://doi.org/10.1037/0096-1523.1.3.280.

Slovic, P. (2007). "If I look at the mass I will never act": Psychic numbing and genocide. *Judgment and Decision Making, 2*(2), 79–95. http://journal.sjdm.org/jdm7303a.pdf.

Small, D., & Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty, 26*(1), 5–16. https://doi.org/10.1023/a:1022299422219.

Small, D. A., Loewenstein, G., & Slovic, P. (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. https://doi.org/10.1016/j.obhdp.2006.01.005.

Smith, R. W., Faro, D., & Burson, K. A. (2013). More for the many: The influence of entitativity on charitable giving. *Journal of Consumer Research, 39*(5), 961–975. https://doi.org/10.1086/666470.

Thomas-Walters, L., & J Raihani, N. (2017). Supporting conservation: The roles of flagship species and identifiable victims. *Conservation Letters, 10*(5), 581–587. https://doi.org/10.1111/conl.12319.

Thornton, B., Kirchner, G., & Jacobs, J. (1991). Influence of a photograph on a charitable appeal: A picture may be worth a thousand words when it has to speak for itself. *Journal of Applied Social Psychology, 21*(6), 433–445. https://doi.org/10.1111/j.1559-1816.1991.tb00529.x.

Tomasello, M. (2020). The moral psychology of obligation. *Behavioral and Brain Sciences, 43*, e56. https://doi.org/10.1017/S0140525X19001742.

Tsuchiya, A., Dolan, P., & Shaw, R. (2003). Measuring people's preferences regarding ageism in health: some methodological issues and some fresh evidence. *Social Science & Medicine, 57*(4), 687–696. https://doi.org/10.1016/S0277-9536(02)00418-5.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95*(3), 371–384. https://doi.org/10.1037/0033-295x.95.3.371.

Wade-Benzoni, K. A., & Tost, L. P. (2009). The egoism and altruism of intergenerational behavior. *Personality and Social Psychology Review, 13*(3), 165–193. https://doi.org/10.1177/1088868309339317.

Van Vugt, M., & Iredale, W. (2013). Men behaving nicely: Public goods as peacock tails. *British Journal of Psychology, 104*(1), 3–13. https://doi.org/10.1111/j.2044-8295.2011.02093.x.

Wang, Y., Tang, Y.-Y., & Wang, J. (2015). Cultural differences in donation decision-making. *PloS one, 10*(9). https://doi.org/10.1371/journal.pone.0138219.

Waytz, A., Iyer, R., Young, L., Haidt, J., & Graham, J. (2019). Ideological differences in the expanse of the moral circle. *Nature Communications, 10*(1), 1–12. https://doi.org/10.1038/s41467-019-12227-0.

Weber, M., Koehler, C., & Schnauber-Stockmann, A. (2019). Why should I help you? Man up! Bystanders' gender stereotypic perceptions of a cyberbullying incident. *Deviant Behavior, 40*(5), 585–601. https://doi.org/10.1080/01639625.2018.1431183.

Weiner, B. (1993). On sin versus sickness: A theory of perceived responsibility and social motivation. *American Psychologist, 48*(9), 957–965. https://doi.org/10.1037/0003-066x.48.9.957.

Wiss, J., Andersson, D., Slovic, P., Västfjäll, D., & Tinghög, G. (2015). The influence of identifiability and singularity in moral decision making. *Judgment and Decision Making, 10*(5), 492–502. http://www.sjdm.org/~baron/journal/13/131003a/jdm131003a.html.

Västfjäll, D., Slovic, P., & Mayorga, M. (2015). Pseudoinefficacy: Negative feelings from children who cannot be helped reduce warm glow for children who can be helped. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00616.

Västfjäll, D., Slovic, P., Mayorga, M., & Peters, E. (2014). Compassion fade: Affect and charity are greatest for a single child in need. *PloS one, 9*(6). https://doi.org/10.1371/journal.pone.0100115.

Zagefka, H., Noor, M., Brown, R., de Moura, G. R., & Hopthrow, T. (2011). Donating to disaster victims: Responses to natural and humanly caused events. *European Journal of Social Psychology, 41*(3), 353–363. https://doi.org/10.1002/ejsp.781.

Zhang, Y., & Slovic, P. (2019). Much ado about nothing: The zero effect in life-saving decisions. *Journal of Behavioral Decision Making, 32*(1), 30–37. https://doi.org/10.1002/bdm.2089