Judgment and Decision Making, Vol. 15, No. 4, July 2020, pp. XX–XX

Moral preferences in helping dilemmas expressed by matching and forced choice

Arvid Erlandsson* Amanda Lindkvist† Kajsa Lundqvist‡ Per A. Andersson§
Stephan Dickert¶ Paul Slovic‖ Daniel Västfjäll**

Abstract

This paper asks whether moral preferences in eight medical dilemmas change as a function of how preferences are expressed, and how people choose when they are faced with two equally attractive help projects. In two large-scale studies, participants first read dilemmas where they "matched" two suggested helping projects (which varied on a single attribute) so that they became equally attractive. They did this by filling in a missing number (e.g., how many male patients must Project M save in order to be equally attractive as Project F which can save 100 female patients). Later, the same participants were asked to choose between the two equally attractive projects. We found robust evidence that people do not choose randomly, but instead tend to choose projects that help female (vs. male), children (vs. adult), innocent (vs. non-innocent), ingroup (vs. outgroup) and existing (vs. future) patients, and imply no (vs. some) risk of a harmful side-effect, even when these projects have been matched as equally attractive as, and save fewer patients than the contrasting project. We also found that some moral preferences are hidden when expressed with matching but apparent when expressed with forced choice. For example, 88–95% of the participants expressed that female and male patients are equally valuable when doing the matching task, but over 80% of them helped female patients in the choice task.

Keywords: moral cognition, expressing moral preferences, helping dilemmas, person trade-offs, prominence effect, medical decision making

1 Introduction

In the movie Sophie's Choice (Pakula, 1979), the main character Sophie is forced by Nazi guards to choose between saving either her daughter's or her son's life. At first Sophie keeps repeating that she cannot choose, implying that her children are exactly equally valuable to her. Still, when reminded that both of her children will be shot if she does not choose, Sophie saves her son (which means giving away her daughter to the guards).

This is first and foremost a dreadful moral dilemma, but it also raises questions about Sophie's preferences regarding her children. Did Sophie really value both her children equally or was it something that made her more prone to save her son when having to make a choice? One could argue that when forced to choose between two exactly equally (un)attractive alternatives, Sophie could (and perhaps should) choose randomly, but this does not seem to be the way most people typically make decisions in the moral realm. On the contrary, although coin-tosses are perceived as fair, they are also perceived as inappropriate when resolving a life-and-death dilemma (Keren & Teigen, 2010).

In this paper we ask: (1) whether people's preferences in a series of moral dilemmas differ when they are expressed in two different ways (matching and forced choice); (2) how people choose when they are faced with two equally attractive alternatives.

1.1 Helping dilemmas

This paper focuses on a specific form of moral dilemma — helping dilemmas (or person trade-offs; Ubel, Richardson & Baron, 2002), which occur when a person learns about two or more need situations where resources are limited, and it is impossible to help everyone in need. In these

This research was financed by a generous grant from the Swedish Science Council (grant number: 2017-01827). We wish to thank the following persons for help during data collection: Anja Grim, Katrine Svane Bech Nielsen, Annica Nilsson, Fanny Plaza, Jacob Andersson, Julia Denkiewicz, Kalle Kallio Strand, Laura Schmitz, Malin Jakobsson Månsson, Justyna Svensson and Agnes Andersson. The results from Study 1 was presented at the Society of Judgment and Decision Making (SJDM) conference in New Orleans in November 2018.

*Linköping University, Department of Behavioral Sciences and Learning. Campus Valla, SE-58183, Linköping, Sweden. E-mail: arvid.erlandsson@liu.se. https://orcid.org/0000-0001-7875-269X.

†Department of Psychology, Göteborg University.

‡Department of Psychology, Lund University.

§Linköping University, Department of Behavioral Sciences and Learning

¶Queen Mary University London and University of Klagenfurt.

‖Decision Research, Eugene, Oregon.

**Linköping University, Department of Behavioral Sciences and Learning and Decision Research, Eugene, Oregon.

situations, people must decide how to allocate help, and in its extreme form this could mean choosing who will live and who will die (e.g., which patient will be connected to the only available respirator). Helping dilemmas are not limited to extraordinary situations or to medical decisions, but happen regularly to most of us. For example, we are responding to a helping dilemma every time we choose to donate to one charity organization but refrain from donating to another (Breeze, 2013; Neumayr & Handy, 2019. Also, for policy makers, politicians, and people working with foreign aid, choosing how to allocate resources between different beneficiaries is a vital part of the job (Alesina & Dollar, 2000; Bucknall, 2003).

In our studies, we asked participants to imagine that they have a job that involves evaluating suggested medical projects designed to help patients. Participants were faced with hypothetical dilemmas each consisting of two suggested helping projects presented next to each other. In each dilemma, participants communicated whether they believed one of the projects was preferable to the other.

1.2 Expressing moral preferences in helping dilemmas

How a person responds to a helping dilemma often reflects her moral preference. For example, a foreign aid official who says YES to most suggested helping projects in Argentina but NO to most (in other ways similar or identical) helping projects in Brazil or Colombia, communicates a moral preference for helping Argentineans.

Moral preferences in helping dilemmas can be expressed in different ways. People can rate the subjective value of each of the helping projects (attractiveness-rating) or distribute resources between the projects (budget-allocation), but in this paper we ask participants to express moral preferences by matching and by forced choice. We focus on these ways because when expressed by the same person, they allow us to ask how people choose when faced with two equally attractive alternatives.

1.2.1 Matching

In economics, indifference curves are often used to demonstrate the value at which two goods give a consumer equal satisfaction and utility. Using the same logic, moral preferences can be expressed by asking participants to "provide their indifference point", or to equate or "match" helping projects so that they become exactly equally attractive (e.g., Ubel et al., 2002). To exemplify, imagine that you hear about a suggested helping Project A that can help Argentinian patients, and an equally costly and otherwise identical Project B that can help Brazilian patients. You learn that Project B will be able to help 100 Brazilians and you are asked how many Argentinians that must be helped in Project A in order to make it equally attractive as Project B. If you respond "100 Argentinians" this implies that you have no preference between helping Argentinians or Brazilians (as helping 100 Argentinians is equally good as helping 100 Brazilians according to you). If you respond "150 Argentinians" this implies that you have a preference for Brazilians (because you need a higher number of Argentinians to make the projects equally good).

1.2.2 Forced choice

The most straightforward way to express a moral preference in a helping dilemma is arguably to choose between Project A and Project B. Importantly, unlike rating, allocation and matching, it is impossible to express indifference between the alternatives when forced to make a choice. Nevertheless, people who have no preference between Argentina and Brazil could choose randomly, for example by throwing a die or flipping a coin (Keren & Teigen, 2010; Dwenger, Kübler & Weizsäcker, 2012). If people really did so, they would be equally likely to end up choosing either of the projects (Shah, Tsuchiya & Wailoo, 2014).[1]

1.3 The Prominence effect: Choosing between two equally attractive helping projects

Matching and forced choice are at the core of a decision making phenomenon called the Prominence Effect: when faced with a trade-off between two or more alternatives, people assign a higher weight to the more important (prominent) attribute when making choices than when rating or matching the alternatives. The prominence effect paradigm began with an unexpected finding by Paul Slovic and led to two influential papers (Slovic, 1975; Tversky, Sattath & Slovic, 1988) where a foundational tenet of decision making – the existence of stable values and preferences – was questioned. Both these papers demonstrated a systematic inconsistency between preferences expressed using a matching task and preferences expressed using a choice task. This inconsistency was explained by an overweighting of the more justifiable attributes when making choices and labeled the Prominence Effect.

For example, in Slovic (1975), participants first matched e.g., pairs of baseball players that differed on two attributes, so that the two players would be equally valuable for their team. This was done by writing how many home-runs Player 1, with a batting average of .287, must hit in a season in order to be equally valuable as Player 2, who had a batting average of .273 and 26 home runs. In a later session, the

[1] A choice task could include a "prefer not to choose" option which makes it possible to express no preference. In these situations it is however not always clear if such a response implies that none of the two helping projects will be materialized (as in Sophie's choice, see e.g., Gordon-Hecker et al., 2017), or if some unspecified other will make the choice (either by their own judgement or randomly).

same participants choose between two players that they had matched to be equally valuable. The results showed that participants did not choose randomly but instead systematically over-selected the player that was superior on the relatively more prominent attribute, which in this case was batting average. To our knowledge, the prominence effect has not yet been documented in the moral domain.

Participants in our studies saw two helping projects that differed on a single attribute (e.g., only ingroup or outgroup patients can be treated) and had one missing piece of information on a scope-related attribute (e.g., the number of outgroup patients possible to treat). First, participants had to match the projects so that they became equally attractive by filling in the missing piece of information. Later, the same participants were forced to make a choice between the two projects that they had rated as equally attractive.

1.4 Number-overriding preferences in helping dilemmas

This paper focuses on *number-overriding preferences in helping dilemmas* which, in our operationalization, occurs anytime a decision maker prefers a helping project that can save fewer over a project that can save more individuals in need. Judgments implying that some lives are valued more than others, or choices that save fewer lives, are expressions of a number-overriding moral preference. Please note that "number" here refers to the number of lives saved.[2]

Neglect of numbers seems to be greater when assessed in separate evaluation (where participants see and respond to only one of the alternatives) than when assessed in joint evaluation helping dilemmas (Hsee & Rottenstreich, 2004; Kogut & Ritov, 2005). The main reason for this is that the number of people that can be helped is much easier to evaluate when presented next to other numbers than when presented in isolation (Hsee & Zhang, 2010). With that said, people are not perfectly number-maximizing even in joint evaluation because the number of people saved is not the only attribute that humans care about. Most people will, e.g., prefer a helping project that can save their own child over a helping project that can save 2, 10 or even 100,000 unknown children. In this paper, we are interested in situational attributes that can elicit number-overriding preferences in helping dilemmas.

[2]Number-maximizing (preferring the option that saves more lives) vs. Number-overriding (preferring the option that saves fewer lives) in helping dilemmas can be linked to characteristically utilitarian vs. non-utilitarian responding. However, as utilitarians typically care not only about the numbers of lives, but also about, e.g., their remaining length and quality, we argue that number-maximization and number-overriding are more accurate descriptions. Number-overriding in our operationalization is thus not necessarily irrational or non-normative.

1.5 Attributes hypothesized to elicit number-overriding preferences and to be prominent in helping dilemmas

Most studies on preferences in medical helping dilemmas have focused on a single varying attribute (e.g., length or quality of lives for patients, e.g., Shah, Tsuchiya & Wailoo, 2014; Ubel et al., 2002; Nord & Johansen, 2014). We here adopt a much broader perspective and investigate eight different attributes typically associated with increased helping, each in a separate dilemma. Besides eliciting number-overriding in helping situations, we suspect that these eight attributes might also be prominent.

1.5.1 Age of victims

People are generally more motivated to help the young than the old (Goodwin & Landy, 2014. One reason for this is that children are assumed to be more innocent and dependent than adults, and young children are rarely held responsible for their own plight (Back & Lips, 1998). Another, more utilitarian, reason is that the anticipated remaining number of quality-adjusted life years (QALY) is higher for every saved child than for every saved adult (Goodwin & Landy, 2014; Bravo Vergel & Sculpher, 2008). Although there are good consequentialist arguments for preferring to save fewer children rather than more adults, this would still count as a number-overriding preference using our operationalization.

1.5.2 Gender of victims

Research indicates that women tend to receive more help than men, and especially so when the helper is a man (e.g., Eagly & Crowley, 1986; Dufwenberg & Muren, 2006; Weber, Koehler & Schnauber-Stockmann, 2019. One reason for this is that helping by males can be used to signal affluence or kindness toward females (van Vugt & Iredale, 2013; Raihani & Smith, 2015). Another reason seems to be rooted in gender-stereotypic perceptions of victim qualities, implying that women are helped more than men because they are seen as sensitive, kind, non-aggressive and ultimately helpless and in need of protection (e.g., Curry, Lee & Rodriguez, 2004). A preference is number-overriding if it favors helping fewer women rather than more men.

1.5.3 Innocence

An attributional account of helping suggests that the perceived causes of the misfortunes of others influence helping. External causes (e.g., bad luck) increase the sense of victim innocence and helping, whereas causes that are seen as internal to and controllable by the person in need elicit anger which make helping less likely (Zagefka et al., 2011; Seacat, Hirschman & Mickelson, 2007; Weiner, 1980). People express feeling less compassion, and has less neural activity

in brain areas associated with emotion, when hearing about victims who had a role in causing their own suffering (Fehse et al., 2015), and representations of victims that typically increase empathic concern tend to backfire when perceived innocence is low (Kogut, 2011). Number-overriding occurs when people prefer to help fewer "innocent" victims rather than more victims that partially caused their own plight.

1.5.4 Group-membership

People are more likely to help members of their ingroup (groups that they identify as members of) than outgroup-members (Dovidio et al., 1997). This tendency has been referred to as intergroup bias, ingroup bias or parochialism (Mackie & Smith, 1998; Baron, 2009) and might arise because of a dislike for the outgroup, liking towards the ingroup (Brewer, 1999), or because people believe that they have a greater responsibility to help ingroup-members (Erlandsson, Björklund & Bäckström, 2015). There exist many types of ingroups, but in this study we focus on nationality, which is arguably one of the most salient naturally occurring but still arbitrary ingroup/outgroup classifications, and meaningful because most people identify themselves as citizens of a country or cultural group (Baron, Ritov & Greene, 2013; Levine & Thompson, 2004; Zagefka, Noor & Brown, 2013). Number-overriding occurs when people prefer to help fewer ingroup-members rather than more outgroup-members.

1.5.5 Existing (vs. future) lives

Intertemporal choices are decisions that involve a trade-off between costs and benefits occurring at different times and the discounted utility model predicts that utilities in the future are discounted by their delay (Samuelson, 1937; Chapman & Elstein, 1995; Bischoff & Hansen, 2016). This intertemporal utility discounting is problematic when people make decisions regarding themselves (should I benefit the existing self or the future self; e.g., retirement savings) but arguably even more problematic when making decisions regarding others (should I help existing others or future others; e.g., Baron & Szymanska, 2011). Discounting may also be one of the major obstacles for combatting climate change as the primary beneficiaries are future generations (Wade-Benzoni & Tost, 2009). Intertemporal discounting can lead to extreme number-overriding meaning that people prefer to help fewer existing victims rather than more future victims.

1.5.6 Avoiding sure death

In helping dilemmas, people often assume that non-helping implies a certain failure (e.g., death) whereas helping implies a certain success (e.g. survival). This is an oversimplification. Instead, estimated outcomes could be better understood by providing an estimated probability for success and failure for those who are helped and those who are not helped, respectively. To illustrate, a medical doctor can estimate the survival-chance to be 10% if a specific patient is not treated in a respirator. The doctor can further estimate that if treated in a respirator, the survival-chance will increase to 40% for the patient. Being treated is no guarantee for survival, but it increases the probability by 30%. We will use the term treatment efficiency to illustrate this increase in survival chance for patients that are treated compared to patients that are not treated.[3]

Research has, however, shown that it is not only the size of the treatment efficiency that matters but also where on the scale (from 0% to 100%) the survival-chance increase occurs. According to the prospect theory (Kahneman & Tversky, 1979; see also Ritov, Baron & Hershey, 1993, and Zhang & Slovic, 2019), we are more sensitive to changes that occur closer to the end points and especially changes that involve 0% (sure death for non-treated patients) and 100% (sure survival for treated patients). This could mean that people prefer projects that can prevent sure deaths (e.g. 0%→40%) or guarantee sure survivals (e.g., 60%→100%) even when these projects are pitted against equally effective projects that can treat more patients but where the survival chances if untreated/treated are located far from the end points (e.g., 30%→70%). This would imply number-overriding.

1.5.7 Avoiding harmful side-effects

Saving 20 people by diverting a runaway trolley onto an empty side-track is the obviously moral thing to do, but saving 50 people by diverting the trolley into a side-track where one person is located seems more problematic despite the net number of lives saved being higher. (Thomson, 1985; Greene, 2008; Bauman et al., 2014). Humans tend to see harm as a result of an action as worse than harm as a result of an omission (Baron & Ritov, 2004), and we are aversive to harmful behavior even if it is an unintended side-effect of an ultimately prosocial act (Anderson, 2003). Likewise, some people are hesitant to use vaccines or airbags that are overall highly beneficial but in rare cases can cause undesired harmful side-effects (Ritov & Baron, 1990). This aversion of incidental harm can lead to number-overriding, meaning that people prefer less efficient helping projects over more efficient helping projects that come with a small risk of a harmful side-effect.

[3]It should also be mentioned that the possible outcomes for helped and non-helped persons are rarely dichotomous but rather comes on a sliding scale (e.g., different degrees of life quality or happiness). The trade-off between the length and the quality of lives is often central in health economy research (e.g. Skedgel, Wailoo & Akehurst, 2015; Nord & Johansen, 2014). Still, in order to not manipulate everything at once, we opted to vary the estimated likelihood of survival but not to differentiate between different levels of life quality in this study.

1.5.8 Causal responsibility

People tend to make appraisals about why a need-situation occurred and consequently who (if anyone) has the causal responsibility and is to be held accountable for the problem at hand (Weiner, 1995). Just as perceived responsibility of the victim can decrease helping, so too can perceived personal responsibility increase it. For example, if a potential helper believes that she caused a specific problem, she is more likely to perceive herself as having a personal responsibility to help victims suffering because of that problem, and to anticipate feeling guilty if she would not help (Erlandsson, Jungstrand & Västfjäll, 2016). Number-overriding occurs when people prefer projects that can help fewer people who are suffering because of a problem they caused over projects that can help more people who are suffering from a problem they did not cause.

1.6 The current studies

In two large-scale studies, we investigate: (1) Which of the eight attributes elicit number-overriding moral preferences when preferences are expressed with matching and choice tasks? (2) Which of the attributes are prominent (relative to the number-attributes)? Recall that the prominent attribute systematically drives preferences among equally attractive options in the choice tasks. If two equally attractive help projects are chosen equally often, then there is no prominent attribute in that dilemma.

1.6.1 Hypothesis

Our initial hypothesis for all eight dilemmas is that most people will express number-overriding preferences in the matching task (albeit to different degrees for different dilemmas), and that they, in the choice-task, will chose the project that is superior on the presumed prominent attribute, even when that project helps fewer people and the two projects have been matched to be equally attractive. If these results materialize, we have initial support for the notion that the prominence effect underlies number-overriding preferences in helping dilemmas.

Several other result patterns are possible. For example, we might find dilemmas where participants express number-overriding preferences in the matching task but chose approximately 50–50 when presented with two projects that they have matched to be equally attractive. We could also find the opposite, namely dilemmas where participants express no moral preference in the matching task but still chose one of the two projects significantly more often in the choice task (i.e., choice-dependent number-overriding).

2 Study 1

Study 1 consisted of two tasks (matching and choice) done by the same participants but temporally separated by at least a month.

2.1 Method

2.1.1 The matching task

Participants. One thousand and seven Swedish participants (401 male, 596 female, 10 unclassified gender, M_{age} = 24.30 years, SD = 6.75) were recruited by 13 research assistants trained to explain the matching task but unaware of the research hypotheses. Participants were approached individually or in small groups at two university campuses during early spring 2017, and (if they agreed to participate) handed a paper-and-pen questionnaire. After completing the questionnaire, participants received a scratch lottery ticket.

Design. We manipulated the questionnaire using a 2×2 design. Half of the participants were randomly assigned to fill in a blank box on the number of treated patients attribute (henceforth Number-condition) whereas the other half filled in a blank box on the treatment-efficiency attribute (average survival-chance for a treated patient; henceforth Efficiency-condition). Further, half of the participants were randomly assigned to fill in the blank on the project presented first in each dilemma (henceforth First-condition), whereas the other half filled in the blank on the project presented second in each dilemma (henceforth Second-condition). See Tables 1 and 2 and the online supplementary material (OSM 1 and 2) for illustrations of the experimental manipulations in both studies.

Procedure and material. Participants were asked to imagine that they had a job where they must make decisions about how to distribute resources between medical projects, and learned that their task, in each dilemma, was to match the two helping projects so that they became exactly equally attractive, by writing a number in the blank box (always shaded in green). "Exactly equally attractive" was explicitly defined as "You would think it was equally good to materialize Project 1 as Project 2".

After reading the instruction page, each participant read and responded to one version of the test dilemma (see Table 1 and OSM 1). The test dilemma was included to let participants familiarize themselves with the layout and to test their comprehension of the matching task. Participants in the Number&First condition filled in how many patients that must be treated in Project 1 (which had a 40% treatment efficiency) for it to be equally attractive as Project 2 (which could treat 100 patients and had a 60% treatment efficiency). A number *lower* than 100 would indicate that the participant

TABLE 1: The test dilemma in all four conditions in the Study 1 matching task. See OSM 1 and 2 for the exact layout of all dilemmas in both studies. (Continued on next page.)

Number&First	**Project 1**	**Project 2**
Who are affected by the disease?	Adults	Adults
Project cost?	400,000SEK	400,000SEK
In which country will the project be implemented?	Sweden	Sweden
Number of ill patients currently in need of treatment?	About 1000 patients currently need treatment	About 1000 patients currently need treatment
What is the average chance of surviving the disease for an ill patient that is NOT treated?	30% chance to survive for each patient that is NOT treated	30% chance to survive for each patient that is NOT treated
What is the average chance of surviving the disease for an ill patient that is treated?	70% chance to survive for each patient that is treated	90% chance to survive for each patient that is treated
Number of patients that will be treated if the project is implemented?	____ ill patients will be treated if the project is implemented	100 ill patients will be treated if the project is implemented

Number&Second	**Project 1**	**Project 2**
Who are affected by the disease?	Adults	Adults
Project cost?	400,000SEK	400,000SEK
In which country will the project be implemented?	Sweden	Sweden
Number of ill patients currently in need of treatment?	About 1000 patients currently need treatment	About 1000 patients currently need treatment
What is the average chance of surviving the disease for an ill patient that is NOT treated?	30% chance to survive for each patient that is NOT treated	30% chance to survive for each patient that is NOT treated
What is the average chance of surviving the disease for an ill patient that is treated?	70% chance to survive for each patient that is treated	90% chance to survive for each patient that is treated
Number of patients that will be treated if the project is implemented?	100 ill patients will be treated if the project is implemented	____ ill patients will be treated if the project is implemented

did not comprehend the matching task (as Project 1 then can treat less patients AND has a lower treatment efficiency than Project 2). Using the same logic, a response *higher* than 100 in the Number&Second condition, a response *lower* than 70% in the Efficiency&First condition, or a number *higher* than 70% in the Efficiency&Second condition) also indicated non-comprehension.[4]

[4]After reading and responding to the test dilemma, participants were asked to explain their response to the research assistant. For participants who responded in a way that indicated non-comprehension on the test dilemma, the assistant spent additional time explaining the matching task before continuing. Participants could change their response for the test dilemma after the explanation, but were not obliged to do so.

After completing the test dilemma, each participant completed the remainder of the questionnaire individually. The layout of all the dilemmas was identical to the test dilemma, but the projects differed so that one attribute was varied at the time (always shaded in orange) while all other attributes were identical in both projects.

The attributes were presented in a fixed order in all dilemmas to make it easier for participants to navigate. A summary of the varying attributes in all dilemmas can be seen in Table 2 (see also OSM 1).

After reading and filling in the blank box in each of the 12 dilemmas, participants could report their gender and age. They were also asked whether they would like to be invited

TABLE 1, CONTINUED.

Efficiency&First	**Project 1**	**Project 2**
Who are affected by the disease?	Adults	Adults
Project cost?	400,000SEK	400,000SEK
In which country will the project be implemented?	Sweden	Sweden
Number of ill patients currently in need of treatment?	About 1000 patients currently need treatment	About 1000 patients currently need treatment
Number of patients that will be treated if the project is implemented?	100 ill patients will be treated if the project is implemented	150 ill patients will be treated if the project is implemented
What is the average chance of surviving the disease for an ill patient that is NOT treated?	30% chance to survive for each patient that is NOT treated	30% chance to survive for each patient that is NOT treated
What is the average chance of surviving the disease for an ill patient that is treated?	____% chance to survive for each patient that is treated	70% chance to survive for each patient that is treated

Efficiency&Second	**Project 1**	**Project 2**
Who are affected by the disease?	Adults	Adults
Project cost?	400,000SEK	400,000SEK
In which country will the project be implemented?	Sweden	Sweden
Number of ill patients currently in need of treatment?	About 1000 patients currently need treatment	About 1000 patients currently need treatment
Number of patients that will be treated if the project is implemented?	100 ill patients will be treated if the project is implemented	150 ill patients will be treated if the project is implemented
What is the average chance of surviving the disease for an ill patient that is NOT treated?	30% chance to survive for each patient that is NOT treated	30% chance to survive for each patient that is NOT treated
What is the average chance of surviving the disease for an ill patient that is treated?	70% chance to survive for each patient that is treated	____% chance to survive for each patient that is treated

to an online follow-up study. Participants who volunteered to participate in the follow-up study wrote their contact information on the last page of the questionnaire.[5]

Inferring moral preferences from the matching task On each dilemma, we converted participants' responses on the matching task to an expression of their moral preference about how to value different lives. For example, in the age dilemma (see Table 2), a participant in condition Number&First who writes a number *higher than* 100 indicates that she thinks that Project A must treat more than 100 adults in order to be equally attractive as Project B which can treat 100 children. This means that she values the life of a child as higher than the life of an adult. Conversely, a participant who writes a number *lower than* 100 indicates that she values the life of an adult as higher than the life of a child, whereas a participant who writes *exactly* 100 indicates that she values children's and adult lives equally high. The opposite is the case for condition Number&Second, and the same logic applies to the Efficiency-conditions.

Excluding participants in the matching task. We removed some participants prior to any analyses using predetermined exclusion criteria. Two participants who failed to respond to five or more of the helping dilemmas, as well

[5]This page was later removed, and participants' contact information was thus linked to their responses in the study only via an ID-number stored separately.

TABLE 2: The four conditions of the dilemmas included in the matching task in Study 1 in the presented order. The letter "X" indicates that participants filled in this value in order to make the two helping projects equally attractive. (Table continued on next page.

Dilemma	Number&First	Number&Second
0. Test dilemma (Comprehension check)	[1] X patients, 30%→70% = [2] 100 patients, 30%→90%	[1] 100 patients, 30%→70% = [2] X patients, 30%→90%
1. Age dilemma	[A] X adult patients = [B] 100 child patients	[A] 100 adult patients = [B] X child patients
2. Gender dilemma	[C] X female patients = [D] 100 male patients	[C] 100 female patients = [D] X male patients
3. Innocence dilemma	[E] X general patients = [F] 100 smoking and drinking patients	[E] 100 general patients = [F] X smoking and drinking patients
4. Comprehension check	[G] X patients for 400,000 SEK = [H] 100 patients for 600,000 SEK	[G] 100 patients for 400,000 SEK = [H] X patients for 600,000 SEK
5. Ingroup dilemma	[I] X Swedish patients = [J] 100 Canadian patients	[I] 100 Swedish patients = [J] X Canadian patients
6. Group size dilemma*	[K] 1000 in need, possible to treat X patients = [L] 200 in need, possible to treat 100 patients	[K] 1000 in need, possible to treat 100 patients = [L] 200 in need, possible to treat X patients
7. Survival chance dilemma 1	[M] X patients, 30%→70% = [N] 100 patients 0%→40%	[M] 100 patients, 30%→70% = [N] X patients 0%→40%
8. Survival chance dilemma 2*	[O] X patients, 30%→70% = [P] 100 patients, 60%→100%	[O] 100 patients, 30%→70% = [P] X patients, 60%→100%
9. Existence dilemma	[Q] X existing patients = [R] 100 future patients	[Q] 100 existing patients = [R] X future patients
10. Personal responsibility dilemma	[S] X patients with a disease you have no connection with = [T] 100 patients with a disease you partially caused	[S] 100 patients with a disease you have no connection with = [T] X patients with a disease you partially caused
11. Attention check	[U] X patients, side effects are headache, cough and running nose = [V] (100 patients, side effects are running nose, cough and headache	[U] 100 patients, side effects are headache, cough and running nose = [V] (X patients, side effects are running nose, cough and headache
12. Side-effect dilemma	[X] X patients, no side effects = [Y] 100 patients, small risk of deadly side effect	[X] 100 patients, no side effects = [Y] X patients, small risk of deadly side effect

as 44 participants who failed the attention check (i.e., did not respond with 100 patients or 70%; see Dilemma 11 in Table 2) were excluded. In addition, participants who responded in a way that indicated misunderstanding of matching task on both comprehension checks (i.e., Dilemmas 0 and 4) were also excluded ($n = 21$). The reported results for the matching task thus include responses from 551 female, 380 male and 9 unclassified participants, $M_{age} = 24.26$ years, $SD = 6.65$.

2.1.2 The choice task

Participants. We prepared choice task invitations for those who had participated in (and not been excluded from) the matching task, and at that time agreed to be contacted again for an online follow-up study (N = 501). Invitations were sent out via e-mail to the address they had provided in the end of the matching task. Participants were offered one or two electronic scratch lottery ticket for participating in the choice task.

A presentation of the study was given directly in the e-mail and the questionnaire was attached as a PDF-file (see OSM 1). Participants responded by replying to the invitation e-mail.

After approximately three weeks of data collection (including three reminders) 151 participants (88 female, 62 male, 1 unclassified, $M_{age} = 24.91$ $SD = 6.84$) had completed the choice task.[6] One participant who failed an attention

[6]The large dropout rate between the two tasks in Study 1 was expected, as participants in the matching task had to actively opt in to even be invited to the choice task (less than 50% did so). Also, the long retention interval between the two tasks likely made participants forget their previous commitment when they received the invitation mail. No remarkable selection

Tabe 2, continued.

Dilemma	Efficiency&First	Efficiency&Second
0. Test dilemma (Comprehension check)	[1] 100 patients, 30%→X% = [2] 150 patients, 30%→70%	[1] 100 patients, 30%→70% = [2] 150 patients, 30%→X%
1. Age dilemma	[A] Adult patients, 30%→X% = [B] Child patients, 30%→70%	[A] Adult patients, 30%→70% = [B] Child patients, 30%→X%
2. Gender dilemma	[C] Female patients, 30%→X% = [D] Male patients, 30%→70%	[C] Female patients, 30%→70% = [D] Male patients, 30%→X%
3. Innocence dilemma	[E] General patients, 30%→X% = [F] Smoking and drinking patients, 30%→70%	[E] General patients, 30%→70% = [F] Smoking and drinking patients, 30%→X%
4. Comprehension check	[G] 400,000 SEK 30%→X% = [H] 600,000 SEK,	[G] 400,000 SEK 30%→70% = [H] 600,000 SEK, 30%→X%
5. Ingroup dilemma	[I] Swedish patients, 30%→X% = [J] Canadian patients, 30%→70%	[I] Swedish patients, 30%→70% = [J] Canadian patients, 30%→X%
6. Group size dilemma*	[K] 1000 in need, 30%→X% = [L] 200 in need, 30%→70%	[K] 1000 in need, 30%→70% = [L] 200 in need, 30%→X%
7. Survival chance dilemma 1	[M] 100 patients, 30%→X% = [N] 100 patients, 0%→40%	[M] 100 patients, 30%→70% = [N] 100 patients, 0%→X%
8. Survival chance dilemma 2*	[O] 100 patients, 30%→X% = [P] 100 patients, 60%→100%	[O] 100 patients, 30%→70% = [P] 100 patients, 60%→X%
9. Existence dilemma	[Q] Existing patients, 30%→X% = [R] Future patients, 30%→70%	[Q] Existing patients, 30%→70% = [R] Future patients, 30%→X%
10. Personal responsibility dilemma	[S] Patients with a disease you have no connection with, 30%→X% = [T] Patients with a disease you partially caused 30%→70%	[S] Patients with a disease you have no connection with, 30%→70% = [T] Patients with a disease you partially caused 30%→X%
11. Attention check	[U] Side effects are headache, cough and running nose, 30%→X% = [V] Side effects are running nose, cough and headache, 30%→70%	[U] Side effects are headache, cough and running nose, 30%→70% = [V] Side effects are running nose, cough and headache, 30%→X%
12. Side-effect dilemma	[X] No side effects, 30%→X% = [Y] Small risk of deadly side effect, 30%→70%	[X] No side effects, 30%→70% = [Y] Small risk of deadly side effect, 30%→X%

Note 1: "30%→70%" illustrate that the average chance of surviving for each untreated patient is 30% whereas the average chance of surviving for each treated patient is 70% (i.e., the treatment efficiency is 40%).

Note 2: Characters in brackets denotes the name of the helping projects as shown to participants.

* For reasons explained in the main text, these dilemmas are not included in this manuscript.

check was excluded prior to any analyses.

Procedure and material. For each invited participant, we created a unique questionnaire including 14 help dilemmas presented similarly to the dilemmas in the matching task. Four of the dilemmas (0, 4, 7 and 12 in Table 4) were identical for all participants and represented manipulation or attention checks. The remaining dilemmas were designed so that the two projects that were pitted against each other had been matched to be exactly equally attractive during the matching task.

Participants' task in each dilemma was to choose which of the two suggested projects to implement. They were asked to choose the project that they found more attractive and in case they found both projects equally attractive they were encouraged to flip a coin, throw a fair die, or use an online number generator to guide them when making the decision. Logically, participants who believed that the two projects were equally attractive would be equally likely to choose either of the two projects (because they would choose at

bias was found when comparing the matching task responses of those who completed vs. did not complete the choice task (see Table 3).

random). If all participants did this, both projects would be chosen approximately equally often on the group level. In contrast, our hypothesis in each dilemma was that the project superior on the presumed prominent attribute would be chosen more often.

2.2 Results

The proportion of participants who, in the matching task, expressed each of the three possible preferences in each condition of each dilemma are presented in Table 3. The number of participants who, in the choice task, chose each of the projects in each dilemma are presented in Table 4.

We coded matching- and choice-task responses so that 1 indicated preferences for the project superior on the presumed prominent attribute, 0 indicated no preference (equal matching) and −1 indicated a preference for the project inferior on the presumed prominent attribute (see Tables 3 and 4).

When aggregating over all eight focus dilemmas, the mean matching-task preference was 0.13 (SD = 0.37), indicating that projects superior on the presumed prominent attributes are preferred when expressing preference with matching, $t(939) = 11.00$, $p < .001$, $d = 0.35$ (one sample t-test with reference value = 0). The mean choice-task preference was 0.59 (SD = 0.32), indicating that, when forced to choose between two equally attractive help projects, people tend to choose the project superior on the prominent attribute, $t(149) = 22.86$, $p < .001$, $d = 1.84$. The rank-order correlation (across participants) between mean matching and choice preferences was $r_s = -.34$. Additional analyses of aggregated preferences and their relation to individual differences are presented in OSM 3.

To increase readability, the results for each dilemma are presented separately in the following text. When reporting the results from the choice task, we report three one proportion z-tests (which compare the actual distribution against a 50–50 distribution) for each dilemma. The first z-test included all participants, the second included only participants who expressed no preference (equal matching) in the matching task, and the third included only participants who expressed that the project superior on the presumed prominent attribute was more attractive.[7] Crucially, the third z-test is our strongest indicator of the prominence effect as it include only participants who choose between two equally attractive projects where one project is superior on the saved lives-attribute and the other project is superior on the presumed prominent attribute.

2.2.1 Age dilemma

Matching task. 44.0% of the participants valued children higher whereas 24.7% valued adults higher (the remaining participants valued children and adults equally high).[8] The mean preference was 0.19 [0.14–0.24] which illustrates a small preference for helping children $t(936) = 7.34$, $p < .001$, $d = 0.23$; one sample t-test with reference value = 0).

Choice task. Project A which helped adult patients was pitted against Project B which helped child patients. Note that in this and all other dilemmas reported below, all participants made a choice between two projects that they had matched to be exactly equally attractive during the matching task. Despite this, the project helping children was chosen by 80.4% of the participants. The Clopper-Pearson 95% confidence interval of the observed proportion was [73.1%–86.5%; Fleiss, Levin & Paik, 2003]. Our first z-test indicated that this was significantly above a 50–50 distribution ($z = 7.39$, $n = 148$, $p < .001$). Second, 81.6% [67.9%–91.2%] of those who had valued adult and children's lives equally in the matching task chose the project that could help children over the project that could help equally many adults ($z = 4.42$, $n = 49$, $p < .001$). Third, 73.0% [61.4%–82.7%] of the participants who had valued children's lives higher than adult lives in the matching task still chose the project that could help fewer children over the project that could help more adults ($z = 3.96$, $n = 74$, $p < .001$).

2.2.2 Gender dilemma

Matching task. A large majority (88.3%; 86.8% of female and 91.0% of male participants) valued female and male patients equally high, whereas only 6.0% [5.8%] valued females [males] higher. The mean preference was 0.00 [−0.02–0.02] implying no preference for helper either gender ($t(938) = 0.19$, $p = .849$, $d < 0.01$). This indicates a clear general preference for valuing female and male patients equally when expressing preferences with a matching task.

Choice task. The project helping female patients was chosen by 84.7% [77.9%–90.1%] of the participants ($z = 8.50$,

[7] This was done because it is conceivable that some participants misunderstood the matching task e.g., by thinking "One adult is 75% as important as one child, so I respond 75 when asked how many adults are equivalent to 100 children" and thus mistakenly expressed a preference in the opposite direction. If the effects in the choice tasks were driven only by these participants, this would severely undermine our arguments. Likewise, expressing no preference in the matching task (equal matching) might result from laziness, a refusal to consider the conflict, or a kind of default response indicating, e.g., "This decision is too hard for me."

[8] In all dilemmas in both studies, we tested if participants' expressed preferences in the matching task differed as a function of which of the four experimental conditions they were in (see Tables 3 and 5). In most dilemmas, participants' preferences did not differ, meaning that the attribute or project they did the matching on, did not change the pattern of preferences. In some dilemmas, preferences did significantly differ as a function of condition but in all but one of these (discussed below), these differences were quantitative rather than qualitative. We therefore aggregated all conditions before the choice-task analyses.

TABLE 3: The proportion of participants in the Study 1 matching task who valued each of the two helping projects higher in each condition in each dilemma, and the mean preference for each dilemma. The projects are presented in the order participants responded to them.

		Number		Efficiency		Total	Preference mean (SD)
Dilemma		First	Second	First	Second		
0. Test	Worse project	15.3%	11.0%	7.0%	7.9%	10.3%	
	Valued equally	6.0%	8.4%	10.0%	12.4%	9.2%	
	Better project	78.7%	80.6%	83.0%	79.8%	80.5%	
1. Age	Adults (−1)	23.6%	28.9%	22.0%	24.1%	24.7% [17.3%]	
	Valued equally (0)	35.8%	31.0%	30.6%	28.3%	31.4% [33.3%]	0.19 (0.81)
	Children (+1)	40.6%	40.2%	47.4%	47.7%	44.0% [49.3%]	
2. Gender	Female patients (+1)	5.2%	7.5%	6.4%	4.6%	6.0% [6.0%]	
	Valued equally (0)	88.7%	86.2%	86.3%	92.0%	88.3% [90.1%]	0.00 (0.34)
	Male patients (-1)	6.1%	6.3%	7.3%	3.4%	5.8% [4.0%]	
3. Innocence	General adults (+1)	32.3%	39.7%	41.5%	43.0%	39.1% [40.4%]	
	Valued equally (0)	29.3%	26.1%	32.9%	28.5%	29.9% [31.1%]	0.08 (0.83)
	Smokers & drinkers (-1)	38.4%	31.2%	25.6%	28.5%	30.9% [28.5%]	
4. Comprehension check	More expensive project	19.8%	11.5%	12.0%	5.6%	12.2%	
	Valued equally	11.5%	11.9%	17.7%	15.1%	14.0%	
	Cheaper project	68.8%	76.6%	70.3%	79.4%	73.8%	
5. Ingroup	Swedish patients (+1)	17.4%	15.9%	18.9%	17.4%	17.4% [22.5%]	
	Valued equally (0)	73.0%	79.5%	73.4%	77.1%	75.8% [72.8%]	0.11 (0.48)
	Canadian patients (-1)	9.6%	4.6%	7.7%	5.5%	6.8% [4.6%]	
6. Group size	Patients from large group (-1)	8.7%	21.4%	57.9%	55.7%	36.1%	
	Valued equally (0)	21.0%	27.7%	27.0%	21.1%	24.2%	
	Patients from small group (+1)	70.3%	50.8%	15.0%	23.2%	39.7%	
7. Survival chance 1	30%-project (-1)	50.7%	56.5%	60.1%	64.1%	57.9% [53.0%]	
	Valued equally (0)	14.0%	15.6%	21.0%	16.5%	16.8% [19.9%]	−0.33 (0.85)
	0%-project (avoid sure death) (+1)	35.4%	27.8%	18.9%	19.4%	25.3% [27.2%]	
8. Survival chance 2	30%-project (-1)	27.0%	32.2%	13.7%	0%	18.2%	
	Valued equally (0)	12.6%	18.8%	20.5%	10.5%	15.6%	
	60%-project (+1)	60.4%	49.0%	65.8%	89.5%	66.2%	
9. Existence	Existing patients (+1)	59.4%	68.2%	73.4%	73.4%	68.7% [72.2%]	
	Valued equally (0)	16.6%	25.1%	17.2%	16.5%	18.9% [22.5%]	0.56 (0.70)
	Future patients (-1)	24.0%	6.7%	9.4%	10.1%	12.5% [5.3%]	
10. Personal responsibility	Unrelated disease (-1)	29.6%	25.7%	24.8%	20.3%	25.1% [23.8%]	
	Valued equally (0)	47.8%	46.0%	46.6%	47.9%	47.1% [45.7%]	0.03 (0.73)
	Disease they caused (+1)	22.6%	28.3%	28.6%	31.8%	27.9% [30.5%]	
11. Attention check	Patients helped by Project U	3.6%	0.8%	3.6%	2.0%	2.5%	
	Valued equally	94.8%	98.0%	94.8%	94.8%	95.6%	
	Patients helped by Project V	1.6%	1.2%	1.6%	3.2%	1.9%	
12. Side-effect	No side-effect (+1)	52.2%	52.5%	68.2%	77.6%	62.7% [66.2%]	
	Valued equally (0)	18.3%	21.8%	18.5%	10.5%	17.3% [19.2%]	0.43 (0.80)
	Risk of side-effect (-1)	29.6%	25.6%	13.3%	11.8%	20.0% [14.6%]	

Note 1. The scores for the eight focus dilemmas do not include the responses from excluded participants. The scores for the comprehension/attention checks (Dilemmas 0, 4 and 11) include responses from all participants.

Note 2. Numbers in parentheses indicate whether the preferred project was superior (+1), or inferior (-1) on the presumed prominent attribute, or if the projects were matched as equally attractive (0).

Note 3. Percentages in brackets illustrate the matching task preferences for those participants who later completed the choice task.

TABLE 4: Number of participants in the Study 1 choice task who chose each project as a function of which project they valued higher in the matching task. The projects are presented in the order participants responded to them.

Dilemma	Project chosen to implement	
0. Test	Worse project [1]	Better project [2]
	1	145
1. Age	Adults [A] (-1)	Children [B] (+1)
Adults (-1)	0	25
Valued equally (0)	9	40
Children (+1)	20	54
2. Gender	Females [C] (+1)	Males [D] (-1)
Male patients (-1)	6	0
Valued equally (0)	114	21
Female patients (+1)	7	3
3. Innocence	General adults [E] (+1)	Smokers & drinkers [F] (-1)
Smokers & drinkers (-1)	40	2
Valued equally (0)	41	5
General adults (+1)	41	18
4. Manipulation check	Better and cheaper [G]	Worse and more expensive [H]
	147	1
5. Ingroup	Ingroup Swedes [I] (+1)	Outgroup Canadians [J] (-1)
Canadian patients (-1)	7	0
Valued equally (0)	101	7
Swedish patients (+1)	26	8
8. Survival chance 1	30%-project [M] (-1)	0%-project (avoid sure death) [N] (+1)
30%-project (-1)	10	69
Valued equally (0)	7	23
0%-project (+1)	14	27
10. Existence	Existing patients [Q] (+1)	Future patients [R] (-1)
Future patients (-1)	8	0
Valued equally (0)	33	0
Existing patients (+1)	81	28
11. Personal responsibility	Unrelated disease [S] (-1)	Disease they caused [T] (+1)
Unrelated disease (-1)	0	36
Valued equally (0)	21	46
Disease they caused (+1)	28	18
12. Manipulation check	Identical Project [U]	Identical Project [V]
	89	57
13. Side-effect	Without side-effect [X] (+1)	With side-effect [Y] (-1)
Risk of side-effect (-1)	21	0
Valued equally (0)	27	2
No side-effect (+1)	58	41

Note 1. Dilemmas 0, 4, and 12 were identical for all participants. In all other dilemmas, participants saw two projects that they previously had matched to be equally attractive.

Note 2. Characters in brackets denote the name of the projects as shown to participants.

Note 3. Numbers in parentheses indicate whether the preferred project was superior (+1), or inferior (-1) on the presumed prominent attribute, or if the projects were equally preferred (0).

Note 4. The group size dilemma (Nr 6) and the Survival chance dilemma 2 (Nr 9) are not included in this manuscript, but included in the raw data file.

Note 5. The attention check (Nr 7) looked like the other dilemmas but the text in the row shaded in orange was: "This is not a real question but a way to test if you are paying attention. Show that you are paying attention by writing [code] in the response box" (see OSM 1).

n = 150, p < .001). Second, among the majority who valued female and male lives equally in the matching task, 84.4% [77.2%–90.1%] chose the project helping females (z = 7.99, n =135, p < .001).[9] Third, 70% [34.75%–93.33] of the few who valued females higher in the matching task, still chose to help fewer females rather than more males (z = 1.27, n =10, p = .206).

2.2.3 Innocence dilemma

Matching task. 39.1% valued "innocent" general patients higher whereas 30.9% valued "non-innocent" smokers & drinkers higher and 29.9% valued innocent and non-innocent patients equally high. The mean preference was 0.08 [0.03–0.14] which illustrates a very small preference for helping innocent patients (t(934) = 3.02, p =.003, d = 0.10).

Choice task. The project helping innocent patients was chosen by 83.0% [75.9%–88.7%] of the participants (z = 8.00, n = 147, p < .001). Second, 89.1% [76.4%–96.4%] of those who had valued innocent patients and non-innocent patients equally in the matching task chose the project helping innocent patients (z = 5.30, n = 46, p < .001). Third, 69.5% [56.1%–80.8%] of those who valued innocent patients higher in the matching task, still chose the project that could help fewer innocent patients over the project that could help more non-innocent patients (z = 3.00, n = 59, p = .003).

2.2.4 Ingroup dilemma

Matching task. A large majority (75.8%) valued ingroup and outgroup patients equally high whereas 17.4% valued ingroup patients higher and 6.8% valued outgroup patients higher. The mean preference was 0.11 [0.07–0.14] which illustrates a small preference for helping ingroup patients (t(937) = 6.72, p <.001, d = 0.23).

Choice task. The project helping ingroup patients was chosen by 89.9% [83.9%–94.2%] of the participants (z = 9.74, n = 149, p < .001). Second, 93.5% [87.1%–97.3%] of those who had valued ingroup and outgroup lives equally in the matching task chose the project helping ingroup patients (z = 9.04, n = 108, p < .001). Third, 76.5% [58.9%–89.3%] of those who had valued ingroup lives higher in the matching task, chose the project that could help fewer ingroup patients over the project that could help more outgroup patients (z = 3.09, n = 34, p = .002).

[9]Among those who had valued lives equally in the matching task, both female (97.4% [90.9%–99.7%], z = 8.26, n = 76, p < .001) and to a lesser extent male participants (67.2% [53.6%–79.0%], z = 2.62, n = 58, p = .009) tended to choose the project that could help female patients over the project that could help equally many male patients.

2.2.5 Patient group size dilemma.

Due to highly varying preferences as a function of which attribute participants did the matching on (number of treated or treatment-efficiency, see Table 3), we have opted to lift this dilemma from this manuscript.[10]

2.2.6 Survival chance dilemma 1

Matching task. 57.9% valued untreated patients with a 30% survival chance higher, whereas 25.3% valued untreated patients with no chance of surviving higher. Against expectations, the mean preference was −0.33 [−0.38 – -.27] which illustrates a small preference for saving untreated patients with a 30% survival chance rather than patients that will unavoidably die if left untreated (t(935) = 11.69, p <.001, d = 0.39).

Choice task. The helping project that could avoid a sure death was chosen by 79.3% [71.9%–85.5%] of the participants, (z = 7.18, n = 150, p < .001). Second, 76.7% [57.8%–90.1%] of those who had valued patients with both diseases equally in the matching task chose the project that could avoid a sure death for treated patients (z = 2.93, n = 30, p = .003). Third, 65.9% [49.5%–80.0%] of those who had valued patients suffering from a disease with a 0% chance of survival if not treated higher in the matching task still chose the project that could help fewer patients (but avoid a sure death for those treated) over the project that could help more patients (z = 2.04, n = 41, p = .042).

2.2.7 Survival chance dilemma 2

During data collection, we realized that this dilemma was flawed. Specifically, it was logically impossible to express a preference for Project M (30% survival chance if untreated) in the Efficiency&Second condition. For this reason, we have lifted this dilemma from this manuscript.

2.2.8 Existence dilemma.

Matching task. 68.7% valued existing patients higher whereas 12.5% valued future patients higher. The mean preference was 0.56 [0.52–0.61] which illustrates a strong preference for helping existing patients (t(937) = 24.43, p <.001, d = 0.80).

[10]The expressed preferences varied substantially in this dilemma, but this was primarily a function of participants' preferences changing as a function of which condition they read χ^2[6, n =937] = 235.17, p < .001. A majority (over 50%) of the participants in the Number-conditions valued patients from the small patient group higher whereas a majority in the Efficiency-conditions valued patients from the large patient-group higher. This finding is interesting (and will be discussed in a separate manuscript) but poses a problem for the choice-task as preferences in this dilemma seem strongly influenced by which dimension one does the matching.

Choice task. The project helping existing patients was chosen by 81.3% [74.1%–87.2%] of the participants (z = 7.67, n = 150, p < .001). Second, everyone (100%) who had valued existing and future victims equally in the matching task, chose the project that could help existing patients. Third, 74.3% [65.0%–82.2%] of those who had valued existing lives higher in the matching task, still chose the project that could help fewer existing patients over the project that could help more future patients (z = 5.07, n = 109, p < .001).

2.2.9 Personal responsibility dilemma

Matching task . 47.1% valued patients suffering from an unrelated disease and patients suffering from a disease that the participant ostensibly caused equally high whereas 27.9% valued patients suffering from the participant-caused disease higher and 25.1% valued patients suffering from an unrelated disease higher. The mean preference was 0.03 [−0.02–0.07] which illustrates no preference for either project, $t(936)$ = 1.17 (p =.243, d = 0.04).

Choice task. The project helping patients with a disease that participants caused was chosen by 67.1% [58.9%–74.6%] of the participants, (z = 4.18, n = 149, p < .001). Second, 68.7% [56.2%–79.5%] of those who had valued patients suffering from the two diseases equally in the matching task chose the project that helped patients suffering from the participant-caused disease (z = 3.06, n = 67, p = .002). However, only 39.1% [25.1%–54.6%] of those who had valued patients suffering from the participant-caused disease higher in the matching task, chose to help fewer patients suffering from the participant-caused disease over the project that could help more patients with the disease that the participant did not cause (z = 1.48, n = 46, p = .139).

2.2.10 Side-effect dilemma

Matching task. 62.7% valued patients that could be treated without any risk for a side-effect whereas 20.0% valued patients that could be treated with a risk for a side-effect higher. The mean preference was 0.43 [0.37–0.48] which illustrates a medium preference for avoiding causing incidental harm ($t(937)$ = 16.25, p <.001, d = 0.54).

Choice task . The no side-effect project was chosen by 71.1% [63.1%–78.2%] of the participants (z = 5.15, n = 149, p < .001). Second, 93.1% [77.2%–99.2%] of those who had valued lives equally in the matching task chose the project without a side-effect (z = 4.64, n = 29, p < .001). Third, 58.6% [48.3%–68.4%] of those who had valued the lives of patients that could be treated without risk of side-effect higher in the matching task, chose to help fewer patients that could be treated without any risk of side-effect over helping more patients that could be treated with a risk of side-effect (z = 1.71, n = 99, p = .087).

2.2.11 Attention and manipulation checks in the choice task.

Dilemma 7 in the choice task was an attention check (one participant failed this) whereas Dilemmas 0 and 4 were manipulation checks (see Table 4). Against expectations, participants did not seem to choose randomly when choosing between two identical helping projects (U and V) in Dilemma 12. Project U was chosen by 61.0% [52.6%–69.0%] of the participants (z = 2.66, n = 146, p = .008). We discuss this finding in the general discussion.

2.3 Summary of Study 1

Our initial hypothesis in all dilemmas was that we would find group-level number-overriding preferences in the matching task, and that people later still would choose the project that was superior on the varying (supposed prominent) attribute disproportionally often, even when that project could save fewer lives. This hypothesis was supported in the Age, Innocence, Ingroup, Existence and Side-effect dilemmas, although the matching task preferences differed much between dilemmas.

The alternative "choice-dependent number-overriding" hypothesis predicted that most people would express no preference in in the matching task, but that they, in the choice task, would prefer the project superior on the supposed prominent attribute rather than choosing at random. We found support for this hypothesis in the Gender dilemma. It also received partial support in the Personal responsibility dilemma. There, participants disproportionally preferred the project that helped patients whose plight they were responsible for when the two projects could help equally many (i.e., had been matched as equally good), but not when the opposing project could help more patients. We did however note that the personal-responsibility manipulation was difficult to convey in a concise way with this paradigm, and we therefore dropped it for Study 2.

In the Survival chance dilemma, we found preferences in opposing directions in the matching and choice tasks. The matching task revealed preferences in the opposite direction of what we predicted in that people valued patients who would die if untreated *less* than patients who had a 30% chance to survive if untreated. In the choice task however, participants responded as predicted and chose the project helping patients who would die if untreated more frequently, even when this meant that less people would be saved.

3 Study 2

Although Study 1 provided strong support for the prominence effect in several of the included helping dilemmas, it suffered from some methodological drawbacks. For example, the fixed order of the dilemmas as well as of the projects in each dilemma could have influenced the results (Ubel et al., 2002; Carney & Banaji, 2012). We preregistered and conducted Study 2 as a well-powered internal replication in order to test the robustness of the obtained results and at the same time controlling for several of the potential problems in Study 1.[11] The presentation as well as the information in in the helping dilemmas were identical or very similar to the ones used in Study 1 (see OSM 1 and 2).

3.1 Method

3.1.1 Participants

Six hundred and five US participants recruited through Amazon Mechanical Turk completed an online questionnaire. Participants were payed $2.

3.1.2 Design

In Study 2, all participants did the matching task on the number of treated patients-attribute (not on treatment-efficiency). In order to control for possible order effects that potentially could have confounded the results in Study 1, the dilemmas in the matching and choice tasks were presented in an order randomized for each participant in Study 2. We also varied the order of two projects, i.e., half of the participants compared Project A (e.g., adults) against Project B (children) whereas the other half compared Project B (children) against Project A (adults). As in Study 1, half of the participants filled in the blank on the project presented first (First-condition) whereas the other half filled in the blank on the project presented last (Second-condition). For each helping dilemma, participants were randomly assigned to read one of four versions of the dilemma (AB/BA × First/Second).

3.1.3 Procedure and material in the matching task

After reading an instruction page, participants saw a tutorial about how to match the two projects. Participants were shown a test dilemma like the one used in Study 1 and read several paragraphs with explanatory text. In the end of the tutorial we asked participants to match the projects in the test dilemma and this response was one of three comprehension checks in the matching task.

After the tutorial, participants read ten dilemmas (in randomized order) and in each they matched the two projects to become equally attractive by writing how many patients must be treated in one project to make it equally attractive as the project it was pitted against (which could always treat 100 ill patients). To make it easier for participants to communicate in case they believed it was impossible to make the two projects equally attractive, we told them to indicate this by writing the number 0 (zero).

Comprehension checks in the matching tasks. Three comprehension checks were included in the matching task (Dilemmas M1–3, see Table 5 and OSM 2). In line with preregistered criteria, we excluded participants who did not comprehend the matching task on two or three of these dilemmas.

3.1.4 Procedure and material in the choice task

Unlike Study 1, participants completed the choice task right after they completed the matching task. We explicitly stated that in case they believed that the two contrasted projects were equally attractive to them, they should choose randomly. To make this alternative even more accessible, we provided participants with an online number generator obtained from www.random.org.

Participants then read 14 dilemmas (randomized order), and in all dilemmas they had to write the name of one of the two proposed helping projects. In 8 of the 14 dilemmas, participants were faced with two helping projects that they previously had matched to be exactly equally attractive. The remaining dilemmas were either comprehension checks, manipulation checks or an attention check (see Table 6).

Attention and comprehension checks in the choice task. We embedded the same attention check as used in Study 1 in the choice task dilemmas. As preregistered, participants who did not pass this check were excluded prior to any analyses.

There were four comprehension checks in the choice task (see Table 6 and OSM 2). As preregistered, we excluded participants who responded in a way indicating that they did not comprehend two or more of these.

In total, we excluded 121 participants for failing the attention check task and an additional 49 participants for missing more than one comprehension check in either the matching task or the choice task. This left us with 435 participants (219 female, 215 male and 1 unclassified, M_{age} = 37.40 years (SD = 10.43) which was more than the 400 we deemed necessary in the preregistration.

3.2 Results

We coded matching- and choice-task responses in the same way as in Study 1 (−1, 0 and +1, see Tables 5 and 6). When aggregating over all seven focus dilemmas, the mean matching-task preference was 0.13 (SD = 0.31), indicating that projects that are superior on the presumed prominent attributes are preferred when expressing preference with

[11] https://osf.io/jrg38/?view_only=107c07abcc054caf97002bc2ed7400ee.

TABLE 5: The proportion of participants in the Study 2 matching task who valued each of the two helping projects higher in each condition, and the mean preference for each dilemma.

Dilemma		AB		BA		Total	Preference mean (SD)
		First	Second	First	Second		
Test Dilemma M1	Worse project		11.9%			11.9%	
	Valued equally		8.6%			8.6%	
	Better project		76.5%			76.5%	
Age dilemma [2.5%]	Adults (-1)	9.6%	19.1%	20.2%	13.7%	15.8%	
	Valued equally (0)	57.7%	60.6%	52.4%	57.8%	56.8%	0.12 (0.65)
	Children (+1)	32.7%	20.2%	27.4%	28.4%	27.4%	
Gender dilemma [0.9%]	Female patients (+1)	2.7%	4.5%	1.9%	3.8%	3.2%	
	Valued equally (0)	97.3%	94.5%	93.2%	94.3%	94.9%	0.01 (0.23)
	Male patients (-1)	0%	0.9%	4.9%	1.9%	1.9%	
Innocence dilemma [4.4%]	Runners & dieters (+1)	24.0%	39.4%	35.7%	28.7%	31.7%	
	Valued equally (0)	45.2%	42.4%	50.0%	47.0%	46.2%	0.10 (0.73)
	Smokers & drinkers (-1)	30.8%	18.2%	14.3%	24.3%	22.1%	
Ingroup dilemma [1.4%]	American patients (+1)	13.0%	12.1%	17.8%	11.5%	12.1%	
	Valued equally (0)	80.0%	82.8%	80.4%	82.4%	81.4%	0.06 (0.43)
	German patients (-1)	7.0%	5.1%	1.9%	12.0%	6.5%	
Survival chance dilemma [7.4%]	30%-project (-1)	56.4%	68.8%	65.7%	44.9%	59.3%	
	Valued equally (0)	13.8%	12.5%	17.2%	18.4%	15.4%	−0.34 (0.86)
	0%-project (avoid sure death) (+1)	29.8%	18.8%	17.2%	36.7%	25.3%	
Existence dilemma [13.8%]	Existing patients (+1)	67.3%	82.5%	90.2%	61.8%	74.9%	
	Valued equally (0)	18.7%	11.3%	8.5%	20.2%	14.9%	0.65 (0.66)
	Future patients (-1)	14.0%	6.2%	1.2%	18.0%	10.1%	
Side-effect dilemma [3.9%]	No side-effect (+1)	47.7%	67.0%	74.5%	44.2%	58.1%	
	Valued equally (0)	32.7%	21.1%	16.3%	32.7%	25.8%	0.42 (0.75)
	Risk of side-effect (-1)	19.6%	11.9%	9.2%	23.1%	16.0%	
Comprehension check M2	More expensive project	20.5%	10.3%	9.7%	22.6%	16.0%	
	Preferred equally	29.5%	20.5%	29.2%	25.8%	26.2%	
	Cheaper project	50.0%	69.2%	61.1%	51.6%	57.8%	
Comprehension check M3	Patients in Project U	6.8%	11.6%	6.0%	10.4%	8.8%	
	Valued equally	90.2%	81.3%	87.3%	81.8%	84.9%	
	Patients in Project V	3.0%	7.1%	6.7%	7.8%	6.3%	

Note 1: The scores for the seven focus dilemmas does not include the responses from excluded participants. The scores for the comprehension checks (M1–3) include responses from all participants.

Note 2. Numbers in parentheses in the first column indicate whether the preferred project was superior (+1), or inferior (-1) on the presumed prominent attribute, or if the projects were matched as equally attractive (0).

Note 3: Percentages in brackets show the number of participants who believed it was impossible to match the projects so that they became equally attractive.

Table 6: Number of participants in the Study 2 choice task who chose each project as a function of which project they valued higher in the matching task.

Dilemma	Project chosen to implement	
Age	Adults [A] (-1)	Children [B] (+1)
Adults (-1)	1	66
Valued equally (0)	29	212
Children (+1)	20	96
Gender	Females [C] (+1)	Males [D] (-1)
Male patients (-1)	8	0
Valued equally (0)	337	72
Female patients (+1)	9	5
Innocence	Runners & dieters [E] (+1)	Smokers & drinkers [F] (-1)
Smokers & drinkers (-1)	90	2
Valued equally (0)	160	32
Runners & dieters (+1)	97	35
Ingroup	Ingroup Americans [I] (+1)	Outgroup Germans [J] (-1)
German patients (-1)	28	0
Valued equally (0)	313	36
American patients (+1)	38	14
Survival chance	30%-project [M] (-1)	0%-project [N] (+1)
30%-project (-1)	123	116
Valued equally (0)	30	32
0%-project (+1)	64	38
Existence	Existing patients [Q] (+1)	Future patients [R] (-1)
Future patients (-1)	37	1
Valued equally (0)	55	1
Existing patients (+1)	231	50
Side-effect	Without side-effect [X] (+1)	With side-effect [Y] (-1)
Risk of side-effect (-1)	67	0
Valued equally (0)	105	3
No side-effect (+1)	144	99
Manipulation check	Identical Project U	Identical Project V
	204	231
Comprehension check C1	Better and cheaper	Worse and more expensive
	546	59
Comprehension check C2	More efficient	Less efficient
	547	58
Comprehension check C3	Treating 100 patients	Treating 3500 patients
	92	513
Comprehension check C4	Treating 2300 patients	Treating 100 patients
	540	65

Note 1: The scores for the eight focus dilemmas do not include the responses from excluded participants. The scores for the comprehension checks (C1–4) include responses from all participants.

Note 2. Characters in brackets denote the name of the projects as shown to participants.

Note 3. Numbers in parentheses indicate whether the preferred project was superior (+1), or inferior (-1) on the presumed prominent attribute, or if the projects were equally preferred (0).

matching ($t(434) = 9.03$, $p < .001$, $d = 0.42$). The mean choice-task preference was 0.56 ($SD = 0.32$), indicating that, when forced to choose between two equally attractive help projects, people tend to choose the project superior on the prominent attribute ($t(434) = 36.34$, $p < .001$, $d = 1.75$). The rank-order correlation across participants between mean matching and choice preferences was $r_s = -.20$. Additional analyses about aggregated preferences and their relation to individual differences are presented in OSM 3.

As in Study 1, we present the results for each dilemma separately. Again, three one-proportion z-tests for each choice-task were conducted to compare the observed proportion against 50–50: (1) Including all participants (2) Including only participants who expressed no preference in the matching task (3) including only participants who expressed a preference for the project superior on the presumed prominent attribute in the matching task.

3.2.1 Age dilemma

Matching. 56.8% valued adult and child patients equally high whereas 27.4% valued children higher and 15.8% valued adults higher (see Table 5). The mean preference was 0.12 [0.05–0.18] which illustrates a small preference for helping children ($t(423) = 3.68$, $p < .001$, $d = 0.18$).

Choice. The results from Study 1 were replicated as 88.2% [CI95: 84.7%–91.1%] chose the project that treated child patients, and this was significantly above a 50–50 distribution ($z = 15.73$, $n = 424$, $p < .001$; see Table 6). Second, 88.0% [83.2%–91.8%] of those who had valued adult and child lives equally high in the matching task chose the project that helped 100 children over the project that helped 100 adults ($z = 11.80$, $n = 241$, $p < .001$). Third, 82.8% [74.7%–89.2%] of those who valued children higher in the matching task chose the project helping fewer children over the project helping more adults ($z = 7.07$, $n = 116$, $p < .001$).

3.2.2 Gender dilemma

Matching. A large majority (94.9%; 94.0% for female and 95.8% for male participants) valued male and female patients equally high in the matching task, whereas only 3.2% [1.9%] valued female [male] patients higher. The mean preference was 0.01 [−0.01–0.04] meaning that participants expressed no preference for helping females over males or vice versa ($t(430) = 1.28$, $p = .201$, $d = 0.04$).

Choice. The results from Study 1 were replicated as 82.1% [78.2%–85.6%] chose the project that treated female patients ($z = 13.33$, $n = 431$, $p < .001$). Second, 82.4% [78.4%–86.0%] of those who had valued male and female lives equally high in the matching task chose the project that treated 100 females over the project that treated 100 males ($z = 13.11$, $n = 409$, $p < .001$).[12] Third, 60% [32.3%–83.66%] of the few who valued female lives higher in the matching task, still chose to help fewer females rather than more males ($z = 0.77$, $n = 15$, $p = .439$).

3.2.3 Innocence dilemma.

Matching. 46.2% valued innocent and non-innocent patients equally high whereas 31.7% valued innocent patients higher and 22.1% valued non-innocent patients higher. The mean preference was 0.10 [0.03–0.17] indicating a small preference for helping innocent patients ($t(415) = 2.69$, $p = .007$, $d = 0.14$).[13]

Choice. The results from Study 1 were replicated as 83.4% [79.5%–86.9%] chose the project that treated innocent patients, ($z = 13.63$, $n = 416$, $p < .001$). Second, 83.3% [77.3%–88.3%] of those who had valued innocent and non-innocent patients equally high in the matching task chose the project that treated 100 innocent patients over the project that treated 100 non-innocent patients ($z = 9.23$, $n = 192$, $p < .001$). Third, 73.5% [65.1%–80.8%] of those who valued innocent patients higher in the matching task chose the project helping fewer innocent patients over the project helping more non-innocent patients ($z = 5.40$, $n = 132$, $p < .001$).

3.2.4 Ingroup dilemma

Matching. A large majority (81.4%) valued ingroup (US) and outgroup (German) patients equally high whereas 12.1% valued ingroup patients higher and 6.5% valued outgroup patients higher. The mean preference was 0.06 [0.02–0.10], indicating a very small preference for helping ingroup patients ($t(428) = 2.70$, $p = .007$, $d = 0.14$).

Choice. The results from Study 1 were replicated, as 88.3% [84.9%–91.2%] chose the project treating ingroup patients ($z = 15.87$, $n = 429$, $p < .001$). Second, 89.7% [86.0%–92.7%] of those who had valued ingroup and outgroup patients equally high in the matching task chose the project that treated 100 ingroup patients over the project that treated 100 outgroup patients ($z = 14.83$, $n = 349$, $p < .001$). Third, 73.1% [59.0%–84.4%] of those who valued ingroup patients higher in the matching task, still chose the project helping fewer ingroup patients over the project helping more outgroup patients ($z = 3.33$, $n = 52$, $p < .001$).

[12] Among those who had valued lives equally in the matching task, both female (91.2% [86.4%–94.7%], $z = 11.77$, $n = 204$, $p < .001$) and to a lesser extent male participants (74.0% [67.4%–79.9%], $z = 6.86$, $n = 204$, $p < .001$) tended to choose the project that could help female patients over the project that could help equally many male patients.

[13] It should be noted that this manipulation was stronger than the innocence manipulation in Study 1. In Study 2, Project E which helped "innocent" patients who exercised regularly and ate nutritious food was pitted against Project F which helped "non-innocent" patients who ate unhealthy, smoke and drank alcohol (see OSM 1 and 2.

3.2.5 Survival chance dilemma

Matching. 59.3% valued patients from the 30→70%-group higher whereas 25.3% valued patients from the 0→40%-group higher. The mean preference was -0.34 [−0.42– −0.26], indicating a small-medium preference for the 30→70% project, $t(402) = -7.97$ p <.001, $d = 0.40$).

Choice. The results from Study 1 were not replicated in the choice task. Only 46.2% [41.3%–51.2%] chose the project that treated patients from the 0→40%-group and this was not significantly different from a 50–50 distribution ($z = 1.53$, $n = 403$, $p = .127$). Likewise, 48.5% [42.0%–55.0%] of those who had valued patients from the 30%→70%-group higher and 51.6% [38.6%–64.5%] of those who had valued patients from the two projects equally high in the matching task, chose the project that treated patients in the 0→40%-group ($z = 0.46$, $n = 239$, $p = .643$ and $z = 0.25$, $n = 62$, $p = .801$ respectively). We address these diverging results in the general discussion.

3.2.6 Existence dilemma

Matching. 74.9% valued existing patients higher whereas 10.1% valued future patients higher in the matching task. The mean preference was 0.65 [0.58–0.71], indicating a large preference for helping existing patients ($t(374) = 19.09$, $p <.001$, $d = 0.98$).

Choice. The results from Study 1 were replicated as 86.1% [82.2%–89.4%] chose the project that treated existing patients ($z = 13.98$, $n = 375$, $p < .001$). Second, 98.2% [90.4%–100%] of those who had valued existing and future patients equally high in the matching task chose the project that treated 100 existing patients over the project that treated 100 future patients ($z = 7.21$, $n = 56$, $p < .001$). Third, 82.2% [77.2%–86.5%] of those who valued existing patients higher in the matching task chose the project helping fewer existing patients over the project helping more future patients ($z = 10.80$, $n = 281$, $p < .001$).

3.2.7 Side-effect dilemma

Matching. 58.1% valued patients that could be treated without any risk of side effect higher whereas 16.0% valued patients that could only be treated with a side effect-risk higher. The mean preference was 0.42 [0.35–0.49] indicating a medium preference for the help project without any side-effect ($t(417) = 11.44$, $p <.001$, $d = 0.56$).

Choice. The results from Study 1 were replicated as 75.6% [71.2%–79.6%] chose the project that could treat patients without any risk of side effect ($z = 10.47$, $n = 418$, $p < .001$). Second, 97.2% [92.1%–99.4%] of those who had valued patients in the without and with side effect projects equally high in the matching task, chose the project that could treat 100 patients without any risk of side effect ($z = 9.81$, $n = 108$, $p < .001$). Third, 59.3% [52.8%–65.5%] of those who valued patients in the no side effect project higher in the matching task chose the project helping fewer patients without risk of side effect over the project helping more patients with a small risk of side effect ($z = 2.90$, $n = 243$, $p = .004$).

3.2.8 Manipulation check

This dilemma, which pitted Project U against identical Project V, is identical to the one used in Study 1 but we here varied the dilemma so that half of the participants read Project U first whereas the other half read Project V first. Unlike Study 1, we found that the two identical projects where chosen equally often (46.9% chose Project U, $z = 1.29$, $n = 435$, $p = .196$).

4 General discussion

This research examined how preferences in moral dilemmas are influenced by the way we express them and how we make choices when faced with two equally attractive helping projects. There are several novel findings to discuss.

For at least five of the included dilemmas (*Age*, *Innocence, Ingroup, Existence* and *Side-effect*), we found the hypothesized pattern of results in both studies. To varying degrees, participants expressed number-overriding moral preferences in the matching task (e.g., they generally matched the projects so that more than 100 future patients had to be helped in order to be equally attractive as a project helping 100 existing patients). In the choice-task however, most of these participants still chose the projects helping existing, young, innocent, ingroup patients without any risk of a harmful side-effect, and they did so even when these projects could help fewer patients and had been matched as equally attractive as the helping project it was pitted against. These results support the notion the prominence effect underlies number-overriding preferences in these helping dilemmas.

A possible mechanism for these results could be the anchoring-and-adjustment heuristic (Tversky & Kahneman, 1974). The number in the cell contrasted against the blank box (e.g., 100 treated patients) could be understood as an anchor, and participants might then adjust their estimates to better fit with their moral preference. However, adjustments tend to be insufficient because people adjust only until reaching the lower boundary of the range of plausible estimates (Epley & Gilovich, 2006). If one's "real" indifference point is closer to the middle of the plausible range, this could explain why people do not choose at random in the choice task.

Two dilemmas (*Gender* and *Ingroup*) clearly stuck out regarding the popularity of equal matches. To illustrate, 88–

95% matched men and women as equally valuable whereas 75–82% did so for outgroup and ingroup patients (all other dilemmas had 14–57% equal matches). However, when these "indifferent" people were later forced to choose, 82–85% helped 100 females rather than 100 males, and 89–94% helped 100 ingroup rather than 100 outgroup patients. These results are consistent with the "choice-dependent number-overriding" hypothesis, and remarkable because they show that participants' preferences in helping dilemmas dramatically change as a function of how we ask them to express their preferences.

Although speculative, our suggested underlying mechanism for choice-dependent number-overriding is people's desire to express justifiable, politically correct, and status-enhancing preferences (Slovic, 1975; Grubbs, et al., 2019). For some dilemmas (e.g., Age, Existence and Side-effect), it is relatively easy and uncontroversial to justify why some lives are valued higher than others. In contrast, the moral preference that all lives are equally valuable becomes especially pronounced when women are pitted against men and as people of different background and ethnicity are pitted against each other. In these situations, expressing anything else than a preference for equal value would likely be seen as socially controversial and possibly upsetting, so most participants express neutrality (equal value) in the matching task. In a choice situation however, there is no way to express these neutral preferences. Then, rather than choosing randomly (which one would do if one truthfully valued men and women and ingroup and outgroup equally), people opt for the alternative that is relatively easier to justify, namely helping women and ingroup patients. This explanation corresponds nicely to research where decisions in a trade-off game changed dramatically when varying which of the options that was framed as the moral one (Capraro & Rand, 2018). Rather than having specific and stable moral preferences, it seems like most people are primarily motivated to express what they think is generally considered to be the most socially accepted moral attitudes.[14]

Importantly, the results obtained in Study 1 and Study 2 correspond well to each other despite being collected in two different samples (undergraduate students in Sweden in Study 1 and MTurk-workers in USA in Study 2), using two different methods (paper and pen in Study 1, online in Study 2), and having different retention intervals between the matching and the choice task (one month later in Study 1, right afterwards in Study 2). This, together with the rather clear effect on most of the dilemmas, suggest that the obtained results are robust and generalizable.

The first take-home message of this article is that people do not choose at random when two equally attractive helping projects are pitted against each other. Instead, they choose the project that is superior on the more prominent attribute, and children, female, innocent, ingroup and existing victims as well as absence of harmful side-effects are all more prominent attributes than the number of individuals possible to save. The second take-home message is that some moral preferences are hidden when it is possible to express indifference but become revealed when we are forced to make a choice.

4.1 Are people ever indifferent in the choice task?

Is it impossible to be indifferent when faced with a forced choice in a helping dilemma? Although we explicitly told participants that they could use a coin flip or a fair die toss (and even provided a random number generator to participants in Study 2), we did not measure how many used it. To further test whether it is possible to be indifferent when making a choice, we included one dilemma where two identical helping projects were pitted against each other. We were initially concerned when Project U was chosen disproportionally more often than identical Project V in Study 1, but upon reflection, this was attributed to a limitation in the methodology. In Study 1, the order of the presented projects in each dilemma was fixed so that Project U was always presented first. According to the "first is best" heuristic (Carney & Banaji, 2012) people tend to prefer the first when faced with two equally attractive alternatives. In Study 2 we varied the order of the helping projects in all dilemmas and as predicted Projects U and V were now chosen equally often.

4.2 Limitations

In this section, we first discuss the dilemma where the results did not turn out as expected, and then consider issues related to the matching task, as well as possible solutions.

4.2.1 The survival chance dilemma

This dilemma stuck out in two ways. First, in line with prospect theory, we expected that helping projects that could avoid a sure death for treated patients would be preferred when pitted against a helping project that could increase the chance of survival quantitatively. In the matching task, we found the opposite pattern in both studies meaning that most participants valued, e.g., 30%→70% patients higher than 0%→40% patients (which would unavoidably die if not treated). One possible explanation for this is a general affect

[14]One could further argue that this ease-of-justification explanation primarily applies for different people in the different dilemmas. Specifically, choosing to help women rather than men seems easier to justify for politically left-leaning people (to compensate for the inherent societal inequalities between men and women), but choosing to help ingroup rather than outgroup patients seems easier to justify for politically right-leaning participants (to avoid signaling disloyalty). Future studies should test this by including individual difference measures such as political orientation and concern for political correctness (Strauts & Blanton, 2015).

heuristic (Slovic et al., 2002) which would predict that participants feel a negative affect towards the project were all untreated patients, and more than half of all treated patients will die. This negative affect elicits a general preference for the opposing project that in comparison seems more promising.

Second, most participants in Study 1 preferred the project that could avoid a certain death in the choice task. Even 65.9% of those who choose between saving more people in the 30%→70% project and fewer people in the 0%→40% project chose the project that could avoid a sure death for treated patients. This was in line with predictions and would suggest that avoiding a sure death is a prominent attribute. To our surprise, Study 2 did not generate the same results and the projects were chosen about equally often. A provisional explanation is that the affect heuristic in the matching task influenced participants in the choice task in Study 2 but not in Study 1. One reason could be the retention interval between matching and choice. The negative affect elicited by the 0%→40% project in the matching task was no longer present when participants in Study 1 completed the choice task a month later, but it could have remained for participants in Study 2 who completed the choice task some minutes later. Another reason could be that the Study 1 choice task tested undergraduate students who responded to our online invitation (presumably more conscientious than those who did not respond) whereas Study 2 tested experienced MTurk-workers (presumably quicker and less deliberate in their responding). Past research has shown that people experiencing time-pressure have different decision making processes than those who do not (e.g., Payne, Bettman & Johnson, 1988).

4.2.2 The matching task

One could argue that expressing preferences using the matching task is counterintuitive as your task is to eliminate your preference rather than expressing it. In addition, the matching task is complex and more cognitively demanding not only compared to the choice task, but also compared to other non-binary ways of expressing preferences such as attractiveness-ratings or budget allocations (see footnote 7). We agree with these concerns and realize that some of the participants might not completely have understood the matching task. At the same time we argue that we did much to mitigate the influence of possible misunderstandings. For example, we had participants justify their responses in a test dilemma (in Study 1), or included a tutorial (in Study 2) to explain the matching task. We also included several comprehension checks and excluded participants who consistently responded in ways which indicated misunderstanding of the matching task. Most importantly, the preference for the project superior on the prominent attribute was, for most dilemmas, found not only among those who matched in ways suggesting misunderstanding or among those who matched equally, but also among participants who matched in the "predicted" direction. This suggests that prominent attributes loom larger in choice than in matching.

Even so, there are alternatives to the traditional matching task. Indifference point can be inferred from several choices presented in quick succession (e.g., Dolan & Tsuchiya, 2011). This would mean asking participants multiple times if they would choose Project A (X adults) or B (100 children) while increasing or decreasing the value of X. (This method may, however, be treated more like the choice task.) A participant's indifference point is the value of X where they switch to the other project. Another alternative way to test how people choose between two equally attractive options could be to simply ask them to imagine that two helping projects are equally attractive to them, and then ask them to choose between these two. Although it remains to be tested, we suspect that we would find the same pattern in the choice task also if we inferred participants indifference points through several choices (at least if there were some time between the two tasks) or if we asked participants to "assume" that the two projects were equally attractive.

Relatedly, one difference between this paper and the seminal paper by Tversky, Sattath and Slovic (1988) is that whereas they had the two options differing on two continuous variables, most of our dilemmas had options which varied on one categorical variable (e.g., children vs. adults or in-group vs. outgroup) and one numerical variable (e.g., number of patients possible to treat), and the matching was always done on the numerical variable. This raises the question of whether the observed effects are results of the content of the presumed prominent attributes or of them being expressed categorically. We note that some categorical attributes can be expressed numerically and that it, in future studies, might be possible to have participants make the matching task also on the presumed prominent attribute (e.g., 50 ten-year old patients = 100 X-year old patients?).

4.3 Conclusion

This study investigated how moral preferences in different medical helping dilemmas change as a function of how preferences are expressed, and how people choose between two equally attractive helping projects. We found that, when faced with two helping projects that had been rated as equally attractive, a significant majority chose the projects that helped children (vs. adult), female (vs. male), innocent (vs. smokers & drinkers), ingroup (vs. outgroup) and existing (vs. future) patients and implied no (vs. some) risk of harmful side-effect. These projects were chosen more often even when they could help fewer patients than the opposing project, and this implies that these attributes influence preferences more when expressed with forced choice than when expressed with matching. This study is the first to

suggest the prominence effect as an underlying mechanism for number-overriding in helping dilemmas, and that some moral preferences that are hidden when it is possible to express indifference become revealed when people are forced to choose.

5 References

Alesina, A., & Dollar, D. (2000). Who gives foreign aid to whom and why? *Journal of Economic Growth, 5*(1), 33–63. http://dx.doi.org/10.1023/A:1009874203400.

Anderson, C. J. (2003). The psychology of doing nothing: Forms of decision avoidance result from reason and emotion. *Psychological Bulletin, 129*(1), 139–166. http://dx.doi.org/10.1037/0033-2909.129.1.139.

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. http://dx.doi.org/10.1016/S0145-2134(98)00098-2.

Baron, J. (2009). Parochialism as a result of cognitive biases. In R. Goodman, D. Jinks, & A. K. Woods (Eds.), *Understanding social action, promoting human rights*. Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195371895.001.0001.

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94*(2), 74–85. http://dx.doi.org/10.1016/j.obhdp.2004.03.003.

Baron, J., Ritov, I., & Greene, J. D. (2013). The duty to support nationalistic policies. *Journal of Behavioral Decision Making, 26*(2), 128–138. http://dx.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). Psychology Press.

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass, 8*(9), 536–554. http://dx.doi.org/10.1111/spc3.12131.

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. http://dx.doi.org/10.1080/15534510.2016.1232204.

Bravo Vergel, Y., & Sculpher, M. (2008). Quality-adjusted life years. *Practical Neurology, 8*(3), 175–182. http://dx.doi.org/10.1136/pn.2007.140186.

Breeze, B. (2013). How donors choose charities: The role of personal taste and experiences in giving decisions. *Voluntary Sector Review, 4*(2), 165–183. http://dx.doi.org/10.1332/204080513x667792.

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

Bucknall, T. (2003). The clinical landscape of critical care: nurses' decision-making. *Journal of Advanced Nursing, 43*(3), 310–319. http://dx.doi.org/10.1046/j.1365-2648.2003.02714.x.

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 & Decision Making, 13*(1). http://journal.sjdm.org/17/171107/jdm171107.html.

Carney, D. R., & Banaji, M. R. (2012). First is best. *PloS one, 7*(6), e35088. http://dx.doi.org/10.1371/journal.pone.0035088.

Chapman, G. B., & Elstein, A. S. (1995). Valuing the future: Temporal discounting of health and money. *Medical Decision Making, 15*(4), 373-386. http://dx.doi.org/10.1177/0272989x9501500408.

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. http://dx.doi.org/10.1177/0011128703256265.

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. http://dx.doi.org/10.1080/00036840903166244.

Dovidio, J. F., Gaertner, S. L., Validzic, A., Matoka, K., Johnson, B., & Frazier, S. (1997). Extending the benefits of recategorization: Evaluations, self-disclosure, and helping. *Journal of Experimental Social Psychology, 33*(4), 401–420. http://dx.doi.org/10.1006/jesp.1997.1327.

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

Dwenger, N. and Kübler, D. F. and Weizsäcker, G. (2012). Flipping a coin: Theory and evidence. Working paper of the Max Planck institute for tax law and public finance No. 2012–14. https://ssrn.com/abstract=2192422.

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.

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. http://dx.doi.org/10.1016/j.obhdp.2014.11.00.

Erlandsson, A., Jungstrand, A., & Västfjäll, D. (2016). Anticipated guilt for not helping and anticipated warm glow for helping are differently impacted by personal responsibility to help. *Frontiers in Psychology, 7*(1475). http://dx.doi.org/10.3389/fpsyg.2016.01475.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science, 17*(4), 311–318. http://dx.doi.org/10.1111/j.1467-9280.2006.01704.x.

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. http://dx.doi.org/10.1080/17470919.2014.980587.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3 ed.). Hoboken: John Wiley & Sons.

Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Vol 3: The neuroscience of morality: Emotion, brain disorders, and development.* (pp. 35–80). MIT Press.

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

Gordon-Hecker, T., Rosensaft-Eshel, D., Pittarello, A., Shalvi, S., & Bereby-Meyer, Y. (2017). Not taking responsibility: Equity trumps efficiency in allocation decisions. *Journal of Experimental Psychology: General, 146*(6), 771–775. http://dx.doi.org/10.1037/xge0000273.

Grubbs, J. B., Warmke, B., Tosi, J., James, A. S., & Campbell, W. K. (2019). Moral grandstanding in public discourse: Status-seeking motives as a potential explanatory mechanism in predicting conflict. *PloS one, 14*(10), e0223749. http://dx.doi.org/10.1371/journal.pone.0223749.

Hsee, C. K., & Rottenstreich, Y. (2004). Music, pandas, and muggers: On the affective psychology of value. *Journal of Experimental Psychology: General, 133*(1), 23–30. http://dx.doi.org/0.1037/0096-3445.133.1.23.

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

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263–291. http://dx.doi.org/10.2307/1914185.

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

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

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

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. http://dx.doi.org/10.3200/socp.144.3.229-245.

Mackie, D. M., & Smith, E. R. (1998). Intergroup relations: Insights from a theoretically integrative approach. *Psychological Review, 105*(3), 499–529. http://dx.doi.org/10.1037/0033-295X.105.3.499.

Neumayr, M., & Handy, F. (2019). Charitable giving: What influences donors' choice among different causes? *Voluntas, 30*(4), 783–799. http://dx.doi.org/10.1007/s11266-017-9843-3.

Nord, E., & Johansen, R. (2014). Concerns for severity in priority setting in health care: A review of trade-off data in preference studies and implications for societal willingness to pay for a QALY. *Health Policy, 116*(2), 281–288. http://dx.doi.org/10.1016/j.healthpol.2014.02.009.

Pakula, A. J. (Director). (1979). Sophies choice [Film]. ITC Entertainment; Universal Pictures.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 534–552. http://dx.doi.org/10.1037/0278-7393.14.3.534.

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

Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making, 3*(4), 263–277. http://dx.doi.org/10.1002/bdm.3960030404.

Ritov, I., Baron, J., & Hershey, J. C. (1993). Framing effects in the evaluation of multiple risk reduction. *Journal of Risk and Uncertainty, 6*(2), 145–159. http://dx.doi.org/10.1007/BF01065355.

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

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. http://dx.doi.org/10.1111/j.1559-1816.2007.00220.x.

Shah, K. K., Tsuchiya, A., & Wailoo, A. J. (2014). Valuing health at the end of life: an empirical study of public preferences. *The European Journal of Health Economics, 15*(4), 389–399. http://dx.doi.org/10.1007/s10198-013-0482-3.

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. http://dx.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. http://dx.doi.org/10.1037/0096-1523.1.3.280.

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment.* (pp. 397–420). Cambridge University Press.

Strauts, E., & Blanton, H. (2015). That's not funny: Instrument validation of the concern for political correctness scale. *Personality and Individual Differences, 80*, 32–40. http://dx.doi.org/10.1016/j.paid.2015.02.012.

Thomson, J. J. (1985). The trolley problem. *Yale Law Journal, 94*(6), 1395–1415.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. http://dx.doi.org/10.1126/science.185.4157.1124.

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

Ubel, P. A., Richardson, J., & Baron, J. (2002). Exploring the role of order effects in person trade-off elicitations. *Health Policy, 61*(2), 189–199. http://dx.doi.org/10.1016/S0168-8510(01)00238-X.

Wade-Benzoni, K. A., & Tost, L. P. (2009). The egoism and altruism of intergenerational behavior. *Personality and Social Psychology Review, 13*(3), 165–193. http://dx.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. http://dx.doi.org/10.1111/j.2044-8295.2011.02093.x.

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. http://dx.doi.org/10.1080/01639625.2018.1431183.

Weiner, B. (1980). A cognitive (attribution)-emotion-action model of motivated behavior: An analysis of judgments of help-giving. *Journal of Personality and Social Psychology, 39*(2), 186–200. http://dx.doi.org/10.1037/0022-3514.39.2.186.

Weiner, B. (1995). Inferences of responsibility and social motivation. In M. P. Zanna (Ed.), *Advances in experimental social psychology*. Academic Press. http://dx.doi.org/10.1016/S0065-2601(08)60402-5.

Zagefka, H., Noor, M., & Brown, R. (2013). Familiarity breeds compassion: Knowledge of disaster areas and willingness to donate money to disaster victims. *Applied Psychology, 62*(4), 640–654. http://dx.doi.org/10.1111/j.1464-0597.2012.00501.x.

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. http://dx.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. http://dx.doi.org/10.1002/bdm.2089.