Moral circles and mind perception shift perceptions of effective altruism

Kyle Fiore Law
University at Albany, State University of New York, kflaw@albany.edu

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd
Part of the Psychology Commons

Recommended Citation
https://scholarsarchive.library.albany.edu/legacy-etd/2725

This Master's Thesis is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive.
Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
MORAL CIRCLES AND MIND PERCEPTION SHIFT PERCEPTIONS OF EFFECTIVE ALTRUISM

by

Kyle F. Law

A Thesis
Submitted to the University at Albany, State University of New York
In Partial Fulfillment of
The Requirements for the Degree of
Master of Arts

College of Arts and Sciences
Department of Psychology
Summer 2021
Abstract

Across a re-analysis of an existing dataset (Study 1; \(N = 96\)) and a higher-powered new study (Study 2; \(N = 300\)), we reveal that moral valuation of environmentalism over humanitarianism predicts less favorable moral judgments of effective altruism (i.e., welfare-maximizing socially distant altruism directed at humans) that is performed at the exclusion of helping animals in need. Furthermore, this relationship is explained by tendencies in mind perception to dehumanize outgroup members and stigmatized humans, rather than tendencies to anthropomorphize animals (Study 2). These findings reveal that granular versus aggregate individual differences in moral circles and mind perception may be worthy of consideration in efforts to increase global welfare through philanthropy.

**Keywords:** effective altruism, moral circles, mind perception, dehumanization, anthropomorphism
Introduction

Over the past 40 years, the world has seen deepening inequality in the distribution of wealth between the rich and the poor. Since 1980, the wealth share controlled by the top 1% has risen from 25% to 40% worldwide (International Monetary Fund, 2020; Killewald et al., 2017; Piketty, 2017; Zucman, 2019). Compounding the severity of the wealth disparity crisis, people living in poorer, developing nations frequently succumb to easily preventable and treatable infectious diseases such as malaria and HIV/AIDS and experience very high infant mortality rates (Dionne, 2017; GiveWell, 2019; Heft-Neal et al., 2018; Singer, 2015) due to a lack of resources. Fortunately, however, advancements in medicine and technology make it increasingly easier for those in affluent societies to meaningfully alleviate global human suffering with fewer resources and less effort (Bloom, 2016; Lecky, 1869; MacAskill, 2018; Pinker, 2012; Singer, 2015). With a mere click of a button, it is possible to save a life on the other side of the world. Nonetheless, in order to effectively reduce human suffering on a global scale, people must first feel moral concern for the suffering of and obligated to help those who are in the most need, people who – from the perspective of those living in the world’s wealthiest nations – are likely geographically and socially distant. In what follows, we will consider across two studies how individual differences in moral concern and mind perception predict differences in moral judgments of welfare-maximizing altruism with pragmatic implications for how these variables ought to be considered in strivings to increase global welfare through altruism. We specifically demonstrate that, when examining judgment and decision-making in the context of altruistic tradeoffs, it is important to take a granular approach to evaluating individual differences in moral concern.
Aggregate and Granular Individual Differences in Moral Concern

The influential notion of the moral circle conceptualizes the strength of moral concern as being represented by a series of concentric circles, with the strongest feeling of moral obligation to ensure the welfare of entities within one’s inner-most circle (typically composed of genetic relatives and close friends) and a weakening feeling of moral obligation for entities as they grow more socially distant and fall within increasingly outward circles (Crimston et al., 2018; Lecky, 1869; Singer 1981; Waytz et al., 2019). The Moral Expansiveness Scale (MES; Crimston et al., 2016) quantifies people’s moral circles by allowing them to indicate the relative moral standing of 30 entities (e.g., a family member, someone from your neighborhood, a cow, an apple tree, a coral reef) by placing them within four different levels, or boundaries, of moral regard: “the inner circle” (which consists of entities deserving the highest level of moral concern), “the outer circle” (which consists of entities deserving moderate moral concern), “the fringes of moral concern” (which consists of entities deserving minimal moral concern) and “outside the moral boundary” (which consists of entities deserving no moral concern whatsoever). The moral standing participants assign to each of the 30 entities is scored on a 3 “inner circle” – 0 “outside the moral boundary” scale and scores are typically aggregated to determine the overall size and expansiveness of participants’ moral circles.

In alignment with this aggregate conceptualization of moral concern, Graham and colleagues (2017) describe how people’s moral circles are shaped by intrapersonal conflicts between two opposing forces: (1) centripetal forces (e.g., familial attachment, ingroup loyalty), and (2) centrifugal forces (e.g., compassion, concerns regarding fairness and equality). Centripetal forces pull inward, promoting smaller moral circles which primarily include socially closer entities (e.g., family members, friends, ingroup members), whereas centrifugal forces push outward, promoting larger moral circles which include more socially distant entities (e.g.,
outgroup members, stigmatized humans, non-human animals, environments) in addition to socially closer entities (Graham et al., 2017).

This model of the moral circle has yielded fruitful insights into how moral concern can linearly contract and expand as a function of people’s moral values (e.g., see Waytz et al., 2019). However, treating the MES strictly as an aggregate measure of the sheer size of an individual’s moral circle may occlude more fine-grained details regarding the relative moral standings of the various entities within an individual’s moral circle. For example, is it always the case that people feel relatively more moral concern for socially distant humans as compared to non-human animals or for socially close humans as compared to socially distant humans? If not, could these more granular individual differences in moral concern predict outcomes in moral and prosocial decision-making that might go unnoticed when moral circles are measured with aggregate scores on the MES?

Emerging research has challenged the assumption that moral concern is always attributed along a single linear dimension, diminishing from socially close humans to socially distant humans to the natural world. For instance, some people assign more moral worth on the MES to nature (e.g., dolphins, rainforests) than outgroup members and stigmatized humans (e.g., Arabs, homosexuals), a pattern that is predicted by individual differences in mind perception (Rottman et al., 2021). Specifically, those who value environmentalism over humanitarianism possess a greater tendency to anthropomorphize animals (i.e., to attribute human mental state qualities to non-human animals) and dehumanize outgroup members and stigmatized humans (i.e., to attribute fewer human mental state qualities to humans). Additionally, the assignment of greater moral value to animals versus humans on the MES predicts favoritism of environmental
versus humanitarian charities in people’s own charitable efforts (Rottman et al., 2021).

**Moral Judgments of Effective Altruism**

Another context in which a granular conceptualization of individual differences in moral concern may offer predictive advantages over an aggregate conceptualization of these differences is in the context of *Effective Altruism*. Effective altruism is a prominent philosophy and growing social movement which posits that donated resources can do precisely the “most good,” or save the greatest number of lives per dollar donated, when donated to the most socially distant others living in extreme poverty (e.g., strangers in severe need living in distant, developing countries; GiveWell, 2019; Singer, 1972, 2015). Effective altruists donate large proportions of their incomes to charities that can (1) alleviate preventable loss of life in the developing world and (2) save lives in the most cost-effective manner. However, a reality endemic to the finite nature of monetary resources is that using them to help socially distant others in severe need often means those same resources cannot be used to help socially closer others in comparatively less need. In other words, in order for people to maximize the utility of their charitable donations, they must set aside their personal relationships and social biases in moral concern to instead prioritize the moral principle of fairness in order to value all life equally when deciding where to donate their resources (Bloom, 2016; Prinz, 2011; Singer, 2015).

These social dilemmas inherent to donation behavior inspired by the effective altruism movement have begun to spur psychological inquiry into how people’s senses of right and wrong equip them to morally evaluate these altruistic tradeoffs in social distance for gains in welfare. Generally speaking, helping behaviors are morally evaluated quite positively, leading to favorable character judgments of those who help others in need (Barasch et al., 2014; Bostyn & Roets, 2016; Carlson & Zaki, 2018; Landy & Uhlmann, 2018; Piazza et al., 2014; Pizarro et al., 2003). However, moral evaluations of helping begin to look very different when people help
socially distant others in severe need instead of socially closer others in comparatively less need. Recent research examining moral judgments of such altruistic tradeoffs in social distance for welfare gains has established that helping socially distant others is perceived as systematically less morally acceptable when it entails not being able to help socially closer others like family members, friends, and community members, negatively affecting moral character judgments and the perceived trustworthiness of donors who wish to maximize welfare with their altruism (Caviola et al., 2021; Everett & Kahane, 2020; Hughes, 2017; Kahane et al., 2018; Law et al., 2021; McManus et al., 2020; Montealegre et al., 2020).

Relatedly, earlier research has demonstrated that people donate more frequently to those who are socially closer and more similar to themselves (Fareri et al., 2012; Goeree et al., 2010; Hoffman et al., 1996), as well as empathize more easily with these individuals (Cikara et al., 2011; Masten et al., 2010; Preston & de Waal, 2002), and assign more moral value to them (Crimston et al., 2016). This moral prioritization of partiality can help to explain why people are often morally dubious of socially distant altruism. However, not all people make the same kinds of moral tradeoffs, thus suggesting potential individual differences in how favorable people might be toward helping geographically or ontologically distant others. While variables like the extent to which people view all of humanity as their ingroup (i.e., The Identification With All of Humanity Scale [IWAHs]; McFarland et al., 2012) have been shown to moderate moral judgments of socially distant altruism (e.g., Law et al., 2021), the influence of granular individual differences between people’s moral circles on these moral judgments has yet to be examined in the published literature.
Do Moral Circles and Mind Perception Shift Moral Judgments of Effective Altruism?

Consider, for example, a scenario in which an affluent person living in the United States has the opportunity to save the lives of eight socially distant humans by donating $7,000 to a humanitarian charity which provides for people living in Africa to be equipped with insecticidal bed nets, protecting them from the deadly, mosquito-borne disease, malaria. In isolation, donating to this cause may seem to be unequivocally morally praiseworthy. Now, separately consider a second scenario in which an affluent person living in the United States has the opportunity to save the life of an injured chimpanzee by donating $7,000 to an environmental charity which will provide for this chimpanzee to be treated and cared for until it can be reintroduced into its natural habitat. Again, in isolation, donating to this cause may seem unequivocally morally praiseworthy. Finally, consider a third scenario in which an individual decides to donate $7,000 to the humanitarian malaria charity instead of the environmental chimpanzee charity. Because this third scenario involves a tradeoff between donating to different categories of potential beneficiaries (i.e., distant humans versus non-human animals), moral perceptions of the donor’s decision might vary from person to person.

One potential explanation for these anticipated differences in moral judgments of the third above scenario could be related to individual differences in moral concern. The extant literature has demonstrated that some people feel more moral concern for animals and environments than they do for outgroup members and stigmatized humans (Rottman et al., 2021), predicting tendencies to dehumanize outgroup members (e.g., people from distant countries), anthropomorphize animals, and preferences toward donating to environmental versus humanitarian charities. Thus, could it be the case that these granular distinctions in moral concern might also predict differences in moral judgments of effective altruism informed
tradeoffs when these tradeoffs involve decisions to donate resources to humanitarian causes that have the potential to save a greater number of lives instead of environmental causes that have the potential to save fewer lives? Furthermore, would these granular distinctions in moral concern explain more about moral judgments than moral concern considered in terms of the overall size of participants’ moral circles? Finally, might mind perception tendencies to dehumanize humans and anthropomorphize animals associated with individual differences in moral circles also play a role in moral judgments?

It should be noted that the effective altruism philosophy does not necessarily prioritize saving human lives over animal lives. In fact, the philosophy at times argues for its subscribers to donate to environmental charities that have been vetted to have a meaningful impact on global welfare (Singer, 2015). However, the philosophy always advocates for donating to charities that can do “more good” (i.e., save more lives) over charities that can do “less good” (i.e., save fewer lives). That being said, there may arise situations in which donating to a humanitarian charity over an environmental charity can do “more good” and others in which donating to an environmental charity over a humanitarian charity can do “more good.” It is in these contexts – when engaging in effective altruism entails not only a tradeoff in the amount of welfare provided between two charitable options, but also a tradeoff in the type of beneficiary provided for between two charitable options – where we hypothesize that granular individual differences in moral concern will predict differences in moral judgments of altruism. If inspiring more people to engage in the type of donation behavior advocated by the effective altruism philosophy is indeed a societal goal, understanding how these granular or relative differences between people’s moral circles relates to perceptions of effective altruism may prove to be an important step in this process.
The Present Research

While extant research has addressed the extent to which mind perception and the assignment of greater moral value to animals and environments as compared to humans predict people’s own charitable intentions, no published study has investigated whether differences in moral circles and mind perception predict third-party moral judgments of other people’s charitable donations in the context of effective altruism. In two cross-sectional studies, we show that granular individual differences in feeling more moral concern for animals and environments – but not moral concern considered in the aggregate – predict less favorable moral judgments of welfare-maximizing socially distant altruism directed at humans performed at the exclusion of helping an animal (i.e., a chimpanzee) in need (Study 1, a reanalysis of previously collected data from Law et al., 2021). In Study 2, we demonstrate that tendencies to dehumanize outgroup members and stigmatized humans, but not tendencies to anthropomorphize animals, account for this relationship. The current research not only replicates previous findings identifying a link between mind perception and moral concern (Rottman et al., 2021), but also extends these findings by demonstrating how these granular individual differences have significant implications for moral perceptions of effective altruism, which may be due pragmatic consideration in efforts to proliferate the effective altruism philosophy and social movement. The data and materials from Studies 1 and 2 are open to the public at the following link: https://osf.io/82bny/?view_only=b738422f6e6d468987116a1bc426f011.

Study 1

Method

Participants. As Study 1 constitutes a reanalysis of previously collected data, the a priori power analysis conducted was not tailored to the particular research question addressed in Study 1. We recruited 118 participants from MTurk in exchange for a small payment, 22 of whom were excluded for failing at least one attention check, for not completing all five repeated
measures, or for not completing the Moral Expansiveness Scale. Thus, our final sample consisted of 96 participants (mean age = 34.5, SD = 10.8, 41.7% female). We conducted a post-hoc sensitivity analysis which revealed we could expect to reliably detect a bivariate relationship with an effect size of $|r| = 0.32$ with 90% power, specifying an alpha of $\alpha = .05$ and a two-tailed test given our final sample size.

**Procedure.** In a within-subjects design, participants in Study 1 read five vignettes in which an actor donated money to a welfare-maximizing, socially distant cause (e.g., saving multiple lives of people in distant countries) as opposed to one of five alternative entities of varying social distance to the actor (i.e., a family member, a friend, a community member, a countryperson, a chimpanzee). The presentation order of the vignettes was randomized between-subjects. For the sake of the current reanalysis, only the data involving the trial in which the actor donated money to a socially distant cause instead of a chimpanzee were analyzed. For results from the other four conditions, see Supplemental Materials.

**Measures.**

*Moral Acceptability of Actor’s Decision.* After reading each vignette, participants reported the moral acceptability of the decision made by the actor in the vignette on a 9-point scale (“To what extent was it morally acceptable for the actor to donate to [socially distant cause] instead of 1 chimpanzee?”, 1= *completely unacceptable* to 9= *completely acceptable*).

*Moral Expansiveness Scale.* Participants’ moral expansiveness was measured using the 30-item Moral Expansiveness Scale (MES; Crimston et al., 2016). The composite sum of moral worth participants assigned to nine animal and environment entities on the MES was subtracted from that which participants assigned to nine stigmatized and outgroup human entities, yielding a *humans minus animals* score. An
aggregate score across all items on the MES was also calculated. See Supplemental Materials for a detailed description of the scoring procedure.

Results

Moral Acceptability of Actor’s Decision on Aggregate MES Scores. As we predicted, the bivariate relationship between aggregate scores on the MES and moral judgments of socially distant altruism was non-significant, $r(94) = -0.14, p = .16$.

Moral Acceptability of Actor’s Decision on Humans Minus Animals. Providing initial evidence that granular individual differences in moral concern are related to moral perceptions of altruistic tradeoffs, a bivariate correlational analysis revealed a moderate, positive correlation between Humans Minus Animals scores and moral judgments of socially distant altruism performed at the exclusion of helping an animal target, $r(94) = 0.43, p < .001$. In other words, participants who assigned more moral value to animals and environments as compared to outgroup members and stigmatized humans found altruism directed at socially distant humans less morally acceptable.

Study 2

While Study 1 demonstrates a link between individual differences in the relative moral standing of humans and non-human targets within participants’ moral circles and their moral perceptions of socially distant altruism, Study 2 attempts to replicate this pattern in a high-powered sample and extend it by investigating the extent to which individual differences in mind perception (i.e., anthropomorphizing and dehumanization) explain the relationship measured in Study 1.

Method

Participants. An a priori power analysis revealed that we would need to collect data from a sample of $N = 259$ to detect a significant correlation coefficient of $|r| = .2$ with 90%
power. We took a conservative approach and decided that we would collect usable data from a target sample size of 300 subjects. We recruited 479 participants from MTurk in exchange for a small payment, 179 of whom were excluded for failing an attention check and an additional 14 of whom were excluded from the primary analyses for missing data on the Moral Expansiveness Scale (Crimston et al., 2016). Thus, our final sample consisted of 286 participants (mean age = 35.2, $SD = 9.98$, 44.8% female). A post-hoc sensitivity analysis revealed we could expect to reliably detect a bivariate relationship with an effect size of $|r| = 0.19$ with 90% power, specifying an alpha of $\alpha = .05$ and a two-tailed test given our final sample size. Study 2 was preregistered through aspredicted.org (https://aspredicted.org/blind.php?x=en3ku5).

**Procedure.** After completing a mind attribution scale (Rottman et al., 2021; see below for details) and the MES (Crimston et al., 2016), participants in Study 2 read five vignettes similar to those presented in Study 1, in which an actor decided to donate money to a welfare-maximizing, socially distant cause (e.g., a cause that saves multiple lives of people in distant countries) instead of a chimpanzee (see Supplemental Materials for the full text of the study). The presentation order of the vignettes was randomized between-subjects. For exploratory purposes, participants were also presented with an alternate version of the same vignettes in which the actor decided to donate to a socially distant cause instead of a fellow countryperson, but results pertaining to this condition have been moved to Supplemental Materials.

**Measures.**

*Measure of Mind Attribution.* Participants’ tendencies to dehumanize outgroup and stigmatized humans and anthropomorphize animals were measured using the Measure of Mind Attribution (Rottman et al., 2021). Participants reported the degree to which they endorsed statements ascribing the capacity for feelings, thoughts and abilities to 10 outgroup and stigmatized human entities (scores were reverse coded and averaged
to yield a measure of dehumanization) and 10 non-human animal entities (scores were averaged to compose a measure of anthropomorphism). See Supplemental Materials for a detailed description of the scoring procedure.

**Moral Expansiveness Scale.** The MES (Crimston et al., 2016) was administered again in Study 2 and responses on this scale were again used to compute a Humans Minus Animals variable and an aggregate score.

**Moral Acceptability of Actor’s Decision.** After reading each vignette, participants reported the moral acceptability of the decision made by the actor in the vignette on the same scale used in Study 1.

**Results**

**Moral Acceptability of Actor’s Decision on Aggregate MES Scores.** As we predicted, the bivariate relationship between aggregate scores on the MES and moral judgments of socially altruism was non-significant $r(284) = .047, p = .430$.

**Simple Linear Regression.** Prior to running a mediation model including dehumanization and anthropomorphizing as predictors, humans minus animals as the mediator and the perceived moral acceptability of the actor’s decision as the outcome, simple linear regression analyses were conducted to test for potential bivariate relationships between the two predictor variables and the outcome variable and between the mediator and the outcome variable (See Figure 1).

**Humans Minus Animals.** As predicted and replicating what was found in Study 1, a bivariate regression revealed a significant positive relationship between Humans Minus Animals scores and moral judgments of socially distant altruism performed at the expense of helping an animal target. In other words, participants who assigned more moral value to outgroup and stigmatized humans as compared to animals and
environments found it more acceptable for a third-party actor to save the lives of multiple
distant human targets instead of a singular animal target.

**Dehumanization.** In line with our predictions, dehumanization was significantly
and negatively related to moral judgments of socially distant altruism. That is, as
participants exhibited a greater tendency to ascribe reduced humanlike properties to
outgroup members and stigmatized humans, they found it less acceptable for a third-party
actor to save the lives of multiple distant human targets instead of a singular animal
target.

**Anthropomorphism.** Anthropomorphism was not significantly related to moral
judgments of socially distant altruism.
Figure 1. Simple linear regressions revealed that humans minus animals, dehumanization, and anthropomorphism predict moral judgments of socially distant altruism performed at the exclusion of helping an alternative animal target. Dehumanization and anthropomorphism were measured using a 9-point scale with higher scores representing greater dehumanization and anthropomorphism. Positive scores on humans minus animals correspond to assigning greater moral value to outgroup and stigmatized humans as compared to animals and environments. Higher scores on moral acceptability represent greater perceived moral acceptability.

Mediation. A mediation model was analyzed including dehumanization and anthropomorphism entered as predictors, humans minus animals entered as the mediator and the perceived moral acceptability of the actor’s decision entered as the outcome variable (See Figure 2 and Table 1). Replicating the effects found in Rottman et al. (2021), both dehumanization and anthropomorphism were significantly and negatively related to humans minus animals in the full model. In other words, as participants exhibited a greater tendency to dehumanize outgroup members and stigmatized humans and anthropomorphize non-human animals, they trended towards assigning more moral value to animals and environments as compared to humans. However, when all of the variables were included in the model, humans minus animals no longer remained a significant predictor of moral judgments. Thus, neither of the two indirect paths was
significant in the mediation model. Despite the fact that neither of the indirect paths was significant, the negative relationship between *dehumanization* and moral judgments remained significant in the full model, whereas the relationship between *anthropomorphism* and moral judgments did not. So, in sum, it seems from the results of the mediation analysis that participants’ tendencies to dehumanize outgroup members and stigmatized humans fully accounted for the relationship between scores on the MES and moral judgments of altruism that was observed when the variables were analyzed in isolation. Because the data in Study 2 were cross-sectional, it could have been the case that the assignment of moral value to humans versus animals causally preceded mind perception tendencies to dehumanize humans and anthropomorphize animals. As such, we ran a separate exploratory mediation model in which *humans minus animals* (MES) was entered as the predictor with *dehumanization* and *anthropomorphism* entered as mediators (See Supplemental Materials).

![Figure 2. Mediation model shows a significant direct path between dehumanization and the perceived moral acceptability of socially distant altruism.](image)

**Figure 2.** Mediation model shows a significant direct path between dehumanization and the perceived moral acceptability of socially distant altruism.
Table 1. Indirect, component, direct, and total effects from mediation model (Study 2).

<table>
<thead>
<tr>
<th>Type</th>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>Lower</th>
<th>Upper</th>
<th>β</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect</td>
<td>Dehumanization ⇒ Humans Minus Animals (MES) ⇒ Moral Acceptability</td>
<td>-0.0356</td>
<td>0.0228</td>
<td>-0.0802</td>
<td>0.0090</td>
<td>-0.033</td>
<td>-1.56</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td>Anthropomorphism ⇒ Humans Minus Animals (MES) ⇒ Moral Acceptability</td>
<td>-0.0296</td>
<td>0.0188</td>
<td>-0.0665</td>
<td>0.0073</td>
<td>-0.036</td>
<td>-1.57</td>
<td>0.116</td>
</tr>
<tr>
<td>Component</td>
<td>Dehumanization ⇒ Humans Minus Animals (MES)</td>
<td>-1.5669</td>
<td>0.2519</td>
<td>-2.0606</td>
<td>-1.0732</td>
<td>-0.364</td>
<td>-6.22</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Humans Minus Animals (MES) ⇒ Moral Acceptability</td>
<td>0.0227</td>
<td>0.0141</td>
<td>-0.0048</td>
<td>0.0503</td>
<td>0.091</td>
<td>1.62</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>Anthropomorphism ⇒ Humans Minus Animals (MES)</td>
<td>-1.3025</td>
<td>0.1939</td>
<td>-1.6826</td>
<td>-0.9225</td>
<td>-0.394</td>
<td>-6.72</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Direct</td>
<td>Dehumanization ⇒ Moral Acceptability</td>
<td>-0.5175</td>
<td>0.0638</td>
<td>-0.6426</td>
<td>-0.3925</td>
<td>-0.484</td>
<td>-8.11</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Anthropomorphism ⇒ Moral Acceptability</td>
<td>-0.0467</td>
<td>0.0496</td>
<td>-0.1439</td>
<td>0.0505</td>
<td>-0.057</td>
<td>-0.94</td>
<td>0.347</td>
</tr>
<tr>
<td>Total</td>
<td>Dehumanization ⇒ Moral Acceptability</td>
<td>-0.5531</td>
<td>0.0603</td>
<td>-0.6712</td>
<td>-0.4350</td>
<td>-0.517</td>
<td>-9.18</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Anthropomorphism ⇒ Moral Acceptability</td>
<td>-0.0763</td>
<td>0.0464</td>
<td>-0.1671</td>
<td>0.0147</td>
<td>-0.093</td>
<td>-1.64</td>
<td>0.100</td>
</tr>
</tbody>
</table>

Note. Confidence intervals computed with method: Standard (Delta method)
Note. Betas are completely standardized effect sizes

General Discussion

Expanding upon existing models of moral circles (e.g., Crimston et al., 2016; Graham et al., 2017; Lecky, 1869; Singer, 1981; Waytz et al., 2019), emerging research suggests that moral regard is not always attributed across a single linear dimension stretching from close humans to distant humans to the natural world. Instead, some people assign greater moral value to animals and environments than they do to outgroup members and stigmatized humans, due in part to greater tendencies to dehumanize outgroup members and stigmatized humans, to anthropomorphize animals, and to prefer donating to environmentally focused charities as opposed to human-focused charities (Rottman et al., 2021; see Amiot et al., 2020 for a related finding). The current research underscores the importance of the role that granular individual
differences play in the relative moral standing people assign to various entities within their moral circles and in their prosocial decision-making. In two studies, we show that a tendency to feel more moral concern for animals and environments than for outgroup members and stigmatized humans predicts more negative moral judgments altruistic tradeoffs made by charitable donors in scenarios derived from the effective altruism philosophy (Study 1), an effect that goes unnoticed when moral concern is considered as a simple aggregate measure of the sheer size of one’s moral circle. Further, in Study 2, we demonstrate that a tendency for those who feel relatively more concern for nature to dehumanize humans, but not their tendency to anthropomorphize animals, accounts for this relationship.

The current studies serve as a theoretical extension of prior research investigating moral judgments of effective altruism (Everett & Kahane, 2020; Hughes, 2017; Kahane et al., 2018; Law et al., 2021; McManus et al., 2020; Montealegre et al., 2020), drawing attention to two additional individual difference variables that modulate such moral judgments. This research highlights yet another social impediment to doing the “most good” that must be faced by those who seek to impartially maximize welfare by way of their donation behavior, presenting a novel challenge to ideas related to effective altruism put forth in psychology, philosophy and philanthropy. If efforts to inspire donation behavior in line with the effective altruism philosophy are to succeed, granular individual differences in moral concern and individual differences in mind perception may be pragmatically important to consider.

Finally, the current findings may also have implications for recent research examining moral judgments of utilitarian decision-making in non-human animal versus human sacrificial dilemmas (Caviola et al., 2019; Caviola et al., 2020; Wilks et al., 2021). While people generally find it morally wrong to sacrifice a small number of human lives to save a greater number of human lives, this emerging work demonstrates that people find it less wrong to sacrifice a small
number of animal lives to save a greater number of animal lives. Further, Caviola and colleagues (2020) found that this effect is partially explained by speciesism, the tendency people have to morally devalue non-human animals simply because they are not human. In light of the current findings which demonstrate that moral circles and mind perception influence moral judgments of impartial altruism, it could also be the case that these variables influence moral judgments of instrumental harm. That is, individuals who assign more moral value to animals versus humans on the MES and those who anthropomorphize animals might find utilitarian decision-making more morally wrong in sacrificial dilemmas involving animals as compared to those who do not. Future research could look to address this possibility.

While the current studies extend knowledge regarding the relationships between individual differences in moral valuation, mind perception and the morality of effective altruism, there are some notable limitations that should be addressed in future research. Most notably, all of the data collected for this article are cross-sectional, thus no causal nor directional claims can be made regarding the relationships between variables. Future work can look to manipulate variables such as dehumanization and moral valuation in order to develop a more complete understanding of these relationships. Further, as Study 1 constitutes a reanalysis of previously collected data, the sample size was not informed by a power analysis relevant to the particular research in question. While the preregistered and power-analysis informed Study 2 makes up for this to some extent, future replication efforts will be needed to further establish the reliability of these effects. Finally, a potentially worthwhile avenue for future research in this area to explore may be to replicate the effects presented in this paper, while including non-human social opportunity cost targets other than a chimpanzee. It is reasonable to assume that there may be differences between effect sizes when the alternative entity whom does not receive aid in the
vignettes is a wild animal versus a domesticated animal, or a high-sentience animal versus a low-sentience animal (Crimston et al., 2016).

**Conclusion**

As the potential scope of altruism increases, it becomes increasingly more important to understand moral perceptions of socially-distant giving and how these moral perceptions are moderated by variation between individuals’ personal moralities and values. The current research further underscores the importance of considering granular versus aggregate individual differences in the moral concern people feel for human and non-human entities in the context of morality and prosociality (Rottman et al., 2021). Moreover, this paper also demonstrates that mind perception tendencies to dehumanize outgroup members and stigmatized humans weigh heavily upon moral judgments of effective altruism informed tradeoffs between donating to humans vs. non-human animals. In addition to offering theoretical insight into the study of morality and altruism, the current research offers pragmatic insights for future research with the goal of expanding impartial altruism to consider.
References


doi:10.1177/0963721411408713


https://doi.org/10.1016/j.tics.2019.11.012


https://doi.org/10.1523/JNEUROSCI.0610-12.2012


https://doi.org/10.1038/s41586-018-0263-3


Kahane, G., Everett, J. A., Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review, 125*(2), 131-164. http://dx.doi.org/10.1037/rev0000093


Supplementary Materials

Studies 1 and 2 Scoring of “Moral Expansiveness Scale”

Participants’ moral expansiveness was measured using the Moral Expansiveness Scale (MES; Crimston et al., 2016). Participants indicated the relative moral standing of 30 entities (a family member, someone from your neighborhood, a cow, an apple tree, coral reef, etc.) by placing these entities within 4 different levels, or boundaries, of moral regard: an inner circle (entities deserving of “the highest level of moral concern and standing”), an outer circle (entities deserving of “moderate moral concern and consideration”), the fringes of moral concern (entities deserving “minimal moral concern and standing”), and outside the moral boundary (entities deserving of “no moral concern or standing”). Following precedent, the four boundaries were scored as follows: inner circle = 3, outer circle = 2, fringes = 1, and outside the boundary = 0.

For the sake of the current reanalysis, the composite sum of moral worth participants assigned to nine animal and environment entities (e.g., a cow, a pig, a bee, a rainforest, a coral reef) on the MES was subtracted from that which participants assigned to nine stigmatized and outgroup human entities (e.g., a homosexual, a person with different religious beliefs, a person with a different political affiliation), yielding a humans minus animals score (for which positive values represent assigning more moral worth to humans, whereas negative values represent assigning more moral worth to animals and environments). An aggregate score across all items on the MES was also calculated.
**Study 1 Results from Additional Experimental Conditions**

**Helping Decision on Judgments of Moral Acceptability.** Socially distant altruism in Study 1 was viewed as fairly morally acceptable ($M = 6.73, SD = 1.72$) across levels of the independent variable of social distance. A repeated measures ANOVA revealed a significant omnibus effect of the social distance to the actor of the entity that did not receive help on participants’ judgments of moral acceptability, $F(4, 384) = 31.13, p < .001, \eta^2_p = .25$. Post-hoc comparisons with Bonferroni adjustments were used to further decompose this omnibus effect. These comparisons revealed that ratings of moral acceptability of an actor helping multiple socially distant humans (more welfare-maximizing, more socially distant causes) were lower when the entity not receiving help was a family member ($M = 5.35, SD = 2.83$) as compared to when the entity not receiving help was a chimpanzee ($M = 7.90, SD = 1.90$), $t(96) = 8.02, p < .001$, a countryperson ($M = 7.25, SD = 2.19$), $t(96) = 7.12, p < .001$, or a community member ($M = 7.23, SD = 2.22$), $t(96) = 6.92, p < .001$. The moral acceptability of helping distant entities was also lower when the entity not receiving help was a friend ($M = 5.94, SD = 2.68$) as compared to when the entity not receiving help was a chimpanzee, $t(96) = 6.30, p < .001$, a countryperson $t(96) = 5.08, p < .001$, or a community member, $t(96) = 4.87, p < .001$. However, not helping family members and friends was judged to be similarly morally acceptable, $t(96) =$
2.71, \( p = .079 \) (see Supplementary Fig. 1).

**Supplementary Figure 1.** Perceived moral acceptability of socially distant human-directed altruism performed in lieu of helping the entities on the x-axis in Study 1. Higher scores represent perceptions of the donation decision as being more morally acceptable.

**Moral Acceptability of the Actor’s Decision on Aggregate MES Scores.** Participants’ moral expansiveness scores considered in the aggregate did not correlate with perceived moral acceptability of socially distant altruistic action in any of the experimental conditions.

**Moral Acceptability of the Actor’s Decision on Humans Minus Animals.** The extent to which participants assigned more moral value to outgroup and stigmatized humans than they did to animals and environments only predicted moral judgments of socially distant altruism in the condition in which the actor in the vignette denied money to a chimpanzee in order to help socially distant human targets. The relationship was such that as participants assigned more moral value to animals and environments as compared to outgroup and stigmatized humans, they
reported less positive moral judgments of socially distant altruism committed in lieu of helping a chimpanzee in need (see Supplementary Table 1).

**Supplementary Table 1.** Participants’ tendencies to assign more moral value to humans versus animals predicted moral judgments of altruism directed towards socially distant humans committed in lieu of helping a chimpanzee in need.

<table>
<thead>
<tr>
<th>Correlation Matrix</th>
<th>Humans Minus Animals</th>
<th>Moral Acceptability (Chimp)</th>
<th>Moral Acceptability (Countryperson)</th>
<th>Moral Acceptability (Community Member)</th>
<th>Moral Acceptability (Friend)</th>
<th>Moral Acceptability (Family Member)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humans Minus Animals</td>
<td>Pearson's r</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Moral Acceptability (Chimp)</td>
<td>Pearson's r</td>
<td>0.429 ***</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>&lt;.001</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Moral Acceptability (Countryperson)</td>
<td>Pearson's r</td>
<td>0.144</td>
<td>0.168</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.161</td>
<td>0.101</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Moral Acceptability (Community Member)</td>
<td>Pearson's r</td>
<td>0.178</td>
<td>0.236 *</td>
<td>0.596 ***</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.084</td>
<td>0.021</td>
<td>&lt;.001</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Moral Acceptability (Friend)</td>
<td>Pearson's r</td>
<td>0.147</td>
<td>0.140</td>
<td>0.504 ***</td>
<td>0.469 ***</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.155</td>
<td>0.176</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>—</td>
</tr>
<tr>
<td>Moral Acceptability (Family Member)</td>
<td>Pearson's r</td>
<td>0.067</td>
<td>0.140</td>
<td>0.500 ***</td>
<td>0.476 ***</td>
<td>0.701 ***</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.521</td>
<td>0.177</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note. * p < .05, ** p < .01, *** p < .001
**Study 2 Scoring of “Measure of Mind Attribution”**

Participants’ tendencies to dehumanize outgroup and stigmatized humans and anthropomorphize animals were measured using the Measure of Mind Attribution (Rottman et al., 2021). Participants reported the degree to which they endorsed statements ascribing the capacity for feelings, thoughts and abilities (e.g., to have a sophisticated appreciation of beauty, to understand what others are feeling) to 10 outgroup and stigmatized human entities (e.g., a chain smoker, a gay man) and 10 non-human animal entities (e.g., a tree frog, a panda) on a 9-point scale (1 = “strongly disagree” to 9 = “strongly agree”). To compose a measure of dehumanization (higher values represent a greater tendency to ascribe fewer humanlike properties to outgroup members and stigmatized humans), scores on items pertaining to outgroup members and stigmatized humans were reverse coded and subsequently averaged together. To compose a measure of anthropomorphism (higher values represent a greater tendency to assign humanlike properties to non-human animals), scores on items pertaining to non-human animals were averaged together.
**Study 2 Countryperson Supplemental**

**Supplementary Table 2.** Participants’ tendencies to dehumanize outgroup members and stigmatized humans, but not their tendencies to anthropomorphize animals, nor scores on the MES predicted moral judgments of altruism directed towards socially distant humans committed in lieu of helping a fellow countryperson in need.

Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Moral Acceptability (Countryperson)</th>
<th>Humans Minus Animals</th>
<th>Dehumanization</th>
<th>Anthropomorphism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moral Acceptability</td>
<td>Pearson's r</td>
<td>—</td>
<td>-0.366 ***</td>
<td>0.047</td>
</tr>
<tr>
<td>(Countryperson)</td>
<td>p-value</td>
<td>0.072</td>
<td>&lt;.001</td>
<td>0.418</td>
</tr>
<tr>
<td>Humans Minus Animals</td>
<td>Pearson's r</td>
<td>—</td>
<td>-0.216 ***</td>
<td>-0.241 ***</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.219</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Dehumanization</td>
<td>Pearson's r</td>
<td>—</td>
<td>—</td>
<td>-0.410 ***</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>—</td>
<td>—</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Anthropomorphism</td>
<td>Pearson's r</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note. * p < .05, ** p < .01, *** p < .001
Study 2 Exploratory Mediation Analysis

Exploratory Mediation Model. Because the data in Study 2 were cross-sectional, it could have been the case that the assignment of moral value to humans versus animals causally preceded mind perception tendencies to dehumanize humans and anthropomorphize animals. As such, we ran a separate exploratory mediation model in which humans minus animals (MES) was entered as the predictor with dehumanization and anthropomorphism entered as mediators. In this model (see Supplementary Figure 2 and Supplementary Table 3), mind perception tendencies to dehumanize humans but not to anthropomorphize animals completely mediated the relationship between scores on the MES and moral judgments of socially distant altruism. Regardless of the causal directionality of the relationship between moral circles and mind perception, these results further underscore that tendencies to dehumanize outgroup members and stigmatized humans predict moral judgments of socially distant altruism above and beyond moral circles and anthropomorphizing.
Supplementary Figure 2. Exploratory mediation model shows that mind-perception tendencies to dehumanize outgroup members and stigmatized humans completely mediate the bivariate relationship between scores on the MES and moral judgments of socially distant altruism.
**Supplementary Table 3. Indirect, component, direct, and total effects from exploratory mediation model (Study 2).**

Indirect and Total Effects

<table>
<thead>
<tr>
<th>Type</th>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>Lower</th>
<th>Upper</th>
<th>β</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect</td>
<td>Humans Minus Animals (MES) ⇒ Dehumanization ⇒ Moral Acceptability</td>
<td>0.025</td>
<td>0.007</td>
<td>0.010</td>
<td>0.040</td>
<td>0.099</td>
<td>3.35</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>Humans Minus Animals (MES) ⇒ Anthropomorphism ⇒ Moral Acceptability</td>
<td>0.004</td>
<td>0.003</td>
<td>-0.003</td>
<td>0.010</td>
<td>0.014</td>
<td>1.04</td>
<td>0.299</td>
</tr>
<tr>
<td>Component</td>
<td>Humans Minus Animals (MES) ⇒ Dehumanization</td>
<td>-0.048</td>
<td>0.014</td>
<td>-0.075</td>
<td>-0.022</td>
<td>-0.208</td>
<td>-3.59</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>Dehumanization ⇒ Moral Acceptability</td>
<td>-0.518</td>
<td>0.056</td>
<td>-0.628</td>
<td>-0.408</td>
<td>-0.478</td>
<td>-9.22</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>Humans Minus Animals (MES) ⇒ Anthropomorphism</td>
<td>-0.075</td>
<td>0.017</td>
<td>-0.109</td>
<td>-0.041</td>
<td>-0.248</td>
<td>-4.34</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>Anthropomorphism ⇒ Moral Acceptability</td>
<td>-0.047</td>
<td>0.044</td>
<td>-0.132</td>
<td>0.039</td>
<td>-0.056</td>
<td>-1.07</td>
<td>0.285</td>
</tr>
<tr>
<td>Direct</td>
<td>Humans Minus Animals (MES) ⇒ Moral Acceptability</td>
<td>0.022</td>
<td>0.013</td>
<td>-0.004</td>
<td>0.049</td>
<td>0.090</td>
<td>1.69</td>
<td>0.092</td>
</tr>
<tr>
<td>Total</td>
<td>Humans Minus Animals (MES) ⇒ Moral Acceptability</td>
<td>0.051</td>
<td>0.014</td>
<td>0.023</td>
<td>0.080</td>
<td>0.206</td>
<td>3.55</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Note. Confidence intervals computed with method: Standard (Delta method)

Note. Betas are completely standardized effect sizes