



**HAL**  
open science

# Fair Social Ordering, Egalitarianism, and Animal Welfare

Marc Fleurbaey, Martin van Der Linden

► **To cite this version:**

Marc Fleurbaey, Martin van Der Linden. Fair Social Ordering, Egalitarianism, and Animal Welfare. American Economic Journal: Microeconomics, 2021, 13 (4), pp.466-491. 10.1257/mic.20190091 . hal-03426174

**HAL Id: hal-03426174**

**<https://hal.science/hal-03426174>**

Submitted on 12 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Fair social ordering, egalitarianism, and animal welfare

Marc Fleurbaey and Martin Van der Linden

November 17, 2019

## Abstract

We study fairness in economies where humans consume one private good and one public good representing the welfare of other species. We show that a social evaluator cannot be egalitarian with respect to humans while always respecting humans unanimous preferences. One solution is to respect unanimous preferences only when it does not lead to a decrease in the welfare of other species. Social preferences satisfying these properties reveals surprising connections between concerns for other species, egalitarianism among humans, and unanimity: The latter two imply a form of dictatorship from humans with the strongest preference for the welfare of other species.

**Keywords:** Welfare Economics, Animal Ethics, Egalitarianism, Efficiency, Fairness.

## 1 Introduction

There is a cultural trend toward granting greater respect to non-human species, as well as a greater interest for animal ethics in philosophy and public advocacy (from [Singer](#),

1975, to [Kymlicka and Donaldson, 2011](#)), and it is time to introduce animals into welfare economics. A pioneering contribution has been made in [Blackorby and Donaldson \(1992\)](#) where the question of the management of population size and possible trade-offs with average animal welfare was examined. While that paper was influential in inspiring more work in population ethics *for humans*, it has not led to the development of a branch of animal ethics in welfare economics.

In this paper, we propose to introduce animal welfare as an argument in human preferences, formally analogous to a public good.<sup>1</sup> The main result of this paper is that, under mild conditions of Pareto efficiency and egalitarianism among humans, the most pro-animal preference in the human population must have a strong influence on social preferences. Specifically, we show that social preferences must be based on a reference preference that is at least as pro-animal as the most pro-animal human in the population. This, in turn, leads to pro-animal social preferences in the presence of *even a single* pro-animal human in the population.

The asymmetry in favor of animals is linked to the fact that, when humans care about animals, egalitarianism among humans is incompatible with the Pareto principle. Restricting the Pareto principle to changes which do not harm animals can be viewed as a mild tilt in favor of animals (this restriction only says that the Pareto principle remains silent when animals lose). Alternatively, it can be viewed as a minimal way of incorporating animals in the relevant population to which the Pareto principle applies. As we show, this mild restriction of the Pareto principle is sufficient to break the symmetry between pro-animal and pro-human preferences.

This result is developed in a framework that follows the fairness approach to social choice ([Fleurbaey and Maniquet, 2011](#)), i.e., only ordinal preferences are taken into account when assessing the situation of individuals who care about their own welfare and the welfare of animals. We however assume that human welfare is measured in a way that is comparable (between humans). The trade-off between the interests of humans

---

<sup>1</sup>Throughout, we often talk of non-human species as *animals* for brevity only. Our analysis is formally independent of the set of non-human species that are taken into account.

and animals is mediated by the ordinal preferences of individuals over combinations of human and animal welfare. This simple framework enables us to develop a clear analysis of the argument for a pro-animal bias in social preferences. We believe that the idea of a pro-animal bias in social preferences extends to a more general framework in which multiple attributes of human and animal situations are kept distinct, and in which no measure of welfare is available apart from the ordinal preferences of humans.

Another feature of our model is that it only accounts for animals through *human* preferences for animal welfare. This arguably makes preferences easier to estimate than if animal preferences had to be elicited. Informational considerations apart, a framework in which only human preferences are taken into account can be viewed as a limiting case. If it is found—as we do—that a strong priority must be given to animal welfare even when animals matter only as much as humans care for them, then one may expect the same conclusion to hold when animal preferences are explicitly accounted for.

We consider this paper a first foray into animal ethics with the tools of fair social choice, and hope that future work will explore this fascinating field further. The paper is structured as follows. Section 2 introduces the framework and the type of social preferences (social ordering functions) studied here. Section 3 presents the clash between the full Pareto principle applied to human preferences (over human and animal welfare) and egalitarianism restricted to human welfare. Section 4 introduces the main result of the paper, showing the importance of the least speciesist individual preferences in shaping the social preferences. Section 5 illustrates how the social preferences characterized in Section 4 can violate the full Pareto principle and presents a way to minimize such violations. Section 6 extends the result of Section 4 to the case of a milder form of egalitarianism among humans. Section 7 further extends by introducing ancillary conditions that force the mild egalitarianism of Section 6 to become as strong as in Section 4. Section 8 concludes and discusses possible refinements of the framework.

## 2 The model and a class of social ordering functions

The set of humans is  $H := \{1, \dots, h\}$ , with  $h \geq 2$ . The **intrinsic welfare** of each human  $i \in H$  is  $x_i \in \mathbb{R}_+$ . The index of welfare for all relevant non-human species is  $x_o \in \mathbb{R}_+$ . In principle,  $x_o$  can represent the welfare of an arbitrary set of non-human species. For brevity, however, we call  $x_o$  the **animal welfare** throughout this paper. Similarly, we often refer to other species as *animals*.

Animal welfare  $x_o$  is consumed as a public good by humans.<sup>2</sup> Each human  $i \in H$  has a preference relation  $R_i$  over the bundles  $(x_o, x_i) \in \mathbb{R}_+^2$ . Altruism between humans is not introduced here, as we focus on the role of feelings toward animals.<sup>3</sup> Preferences are orderings, i.e., transitive and complete binary relations. A preference profile  $\mathbf{R} := (R_1, \dots, R_h)$  is a list of preferences for all humans in  $H$ .

A particular preference of interest for this paper is the animal-first lexicographic preference, denoted  $R^{AL}$  and defined as follows: For all  $(x_o, x_i), (y_o, y_i) \in \mathbb{R}_+^2$ ,  $(x_o, x_i) R^{AL} (y_o, y_i)$  if and only if either  $x_o > y_o$ , or  $x_o = y_o$  and  $x_i \geq y_i$ . Such preference check animal welfare first, and human welfare second, giving absolute priority to the former over the latter.

An **allocation** is a vector  $\mathbf{x} := (x_o; x_1, \dots, x_h) \in \mathbb{R}_+^{h+1}$ . A **social ordering** is an ordering of all the allocations in  $\mathbb{R}_+^{h+1}$ . Preference  $R_i$  over bundles  $(x_o, x_i) \in \mathbb{R}_+^2$  extends naturally to a preference over allocations in  $\mathbb{R}_+^{h+1}$ , with  $\mathbf{x} R_i \mathbf{y}$  for any two allocations  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$  if and only if  $(x_o, x_i) R_i (y_o, y_i)$ .

A typical domain of preference profiles is denoted  $\mathcal{D}$ . Let  $\mathcal{R}$  be the set of all profiles containing preferences that are continuous and strictly monotonic in  $x_o$  and  $x_i$ . While our main characterization results are valid for this domain, they also hold when the

---

<sup>2</sup>The fairness literature on public goods includes [Moulin \(1987\)](#) and [Maniquet and Sprumont \(2004, 2005\)](#); see also [Fleurbaey and Maniquet \(2011, Chap. 8\)](#). The key difference between this literature and our paper is that we weaken the Pareto axiom instead of weakening the egalitarian axiom.

<sup>3</sup>For a study of other-regarding preferences between humans in a similar approach, see [Decerf and Van der Linden \(2015\)](#) and [Treibich \(2019\)](#).

domain of profiles allows for  $R^{AL}$  in addition to continuous and monotonic preferences. This enables us to state a few corollaries that involve  $R^{AL}$ . The subdomain of  $\mathcal{R}$  for which all individual preferences are differentiable is denoted  $\mathcal{R}^\partial$ . For a differentiable preference ordering  $R_i$ , the marginal rate of substitution of  $R_i$  at bundle  $(x_o, x_i)$  is denoted by  $MRS(R_i; (x_o, x_i))$ .

A **social ordering function** (SOF)  $\succeq$  is a function that associates every preference profile  $\mathbf{R}$  in some domain  $\mathcal{D}$  with a social ordering  $\succeq_{\mathbf{R}}$ .

For any preference  $R$  on  $\mathbb{R}_+^2$ , let  $\succeq_R^{lex}$  be the social ordering obtained by leximin extension of  $R$ . That is, for any two allocations  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{h+1}$ ,  $\mathbf{x} \succeq_R^{lex} \mathbf{y}$  if, according to  $R$ , the human who receives the worst bundle in  $x$  receives a bundle that is no worse than the human who receives the worst bundle in  $y$ . If according to  $R$ , the human who receives the worst bundle in  $x$  receives a bundle that is indifferent in  $y$ , then  $\mathbf{x} \succeq_R^{lex} \mathbf{y}$  if, according to  $R$ , the human who receives the *second-worst* bundle in  $x$  receives a bundle that is no worse than the human who receives the *second-worst* bundle in  $y$ , and so on. An SOF  $\succeq$  is a **leximin SOF** if, for every profile  $\mathbf{R} \in \mathcal{D}$ , there exists a **reference preference**  $R(\mathbf{R})$  such that  $\succeq_{\mathbf{R}} = \succeq_{R(\mathbf{R})}^{lex}$ . If  $\succeq$  is a leximin SOF and  $(R(\mathbf{R}))_{\mathbf{R} \in \mathcal{D}}$  is the list of reference preferences associated with  $\succeq$ , we say that  $\succeq$  is **based on**  $(R(\mathbf{R}))_{\mathbf{R} \in \mathcal{D}}$ .

Three features of our framework deserve comments and explanations. First, there is no formal difference between  $x_o$  and any other public good humans may want to consume. However, we stress that some of the axioms we study — and in particular our restriction of the Pareto axiom — find natural justifications when  $x_o$  is interpreted as animal welfare, which would not apply if  $x_o$  was another arbitrary public good. It is in this sense that our analysis and results apply specifically to animal ethics. We welcome the possibility to apply our analysis to other public goods that have a special intrinsic value but the case of animals seems especially fitting for it, which is why it is the focus of our discussion. At the same time, we warn against re-interpreting our results with public goods for which our axioms might not be equally justified.

Second, we restrict attention to non-negative numbers for human welfare  $x_i$  and

animal welfare  $x_o$ . This can be interpreted as meaning that we do not consider allocations in which the lives of these beings are not worth living. Our results do not depend on this restriction, but we do want to avoid a situation in which the results would critically depend on allowing “bad” lives to be part of the picture, which is why the framework is limited to “good” lives.

Third, the animal welfare  $x_o$  is a single number, which is a great simplification of the multidimensional reality of complex ecosystems. This simplification is tantamount to assuming that all human beings make the same evaluation of the situation of non-human ecosystems, and that their preferences differ only in the assessment of trade-offs between humans and animals. The number  $x_o$  is then a proxy for their common utility relative to animal welfare. Our negative (impossibility) results do not depend on this simplification but our positive results do depend on it.<sup>4</sup> We discuss how to extend our results to the more complex multidimensional case in the last section of the paper.

### 3 The impossibility of a speciesist Paretian egalitarian

From a speciesist point of view, the interests of humans must always prevail over those of animals. It may be hard for a speciesist social evaluator to reconcile human preferences when they differ. But the preference of a speciesist evaluator should at least coincide with *unanimous* preferences among humans, even when humans unanimously agree to drastically reduce animal welfare. This speciesist respect for unanimous human preferences is captured in the following axiom.

**SPECIESIST PARETO.** For all  $\mathbf{R} \in \mathcal{D}$ , and all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$ , if  $\mathbf{x} R_i \mathbf{y}$  for all  $i \in H$ ,

---

<sup>4</sup>In particular, with a multidimensional description of ecosystems, it would be harder to compare individual preferences in terms of marginal rate of substitution between human and animal welfare, and comparisons of this kind are important to the main result of this paper.

then  $\mathbf{x} \succeq_{\mathbf{R}} \mathbf{y}$ ; if  $\mathbf{x} P_i \mathbf{y}$  for all  $i \in H$ , then  $\mathbf{x} \succ_{\mathbf{R}} \mathbf{y}$ .

This axiom is slightly weaker than the usual Strong Pareto axiom because it does not require a strict social preference when only a subset of individuals have a strict preference and the others are indifferent.

If a speciesist social evaluator is also egalitarian, she must support some form of redistribution between humans with high intrinsic welfare and humans with low intrinsic welfare. To guarantee that inequalities are reduced, it is important that redistributions be performed at a fixed level of animal welfare, as changes in animal welfare may otherwise affect the welfare of humans in a way that more than compensated the redistribution. This is formalized in the next axiom, which requires that a (Pigou-Dalton) transfer be a weak social improvement provided that the level of animal welfare is fixed.

**TRANSFER.** For all  $\mathbf{R} \in \mathcal{D}$  and all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$ , if  $x_o = y_o$ , if there exists  $\Delta \in \mathbb{R}_{++}$  and  $j, k \in H$  such that

$$x_j = y_j - \Delta \geq y_k + \Delta = x_k,$$

and if  $x_i = y_i$  for all  $i \neq j, k$ , then  $\mathbf{x} \succeq_{\mathbf{R}} \mathbf{y}$ .

**Transfer** is a strong egalitarian axiom in the sense that it does not take preferences into account. In particular, it allows for a transfer between humans the indifference curves of whom cross at another level of animal welfare, as illustrated in Figure 1. In other words, **Transfer** may imply that, starting from allocation  $\mathbf{x}$ , a transfer from  $j$  to  $k$  is a social improvement although for some other allocation  $\mathbf{y}$  that leaves  $j$  and  $k$  almost indifferent with  $\mathbf{x}$ , the reversed transfer from  $k$  to  $j$  is a social improvement (see Figure 1). As is well-known (Fleurbaey and Trannoy, 2003), these reversals typically generate impossibilities, and we have the following result.

**Proposition 1.** *On domain  $\mathcal{R}$ , no SOF satisfies both **Speciesist Pareto** and **Transfer**.*

*Proof.* The proof follows the standard strategy presented in Fleurbaey and Maniquet (2011, Chap. 5) and is illustrated in Figure 1. Consider the allocations and preferences represented in the figure, and suppose that  $\succeq$  satisfies both axioms. By **Transfer**



$a \succ_{\mathcal{R}} b$ . By **Speciesist Pareto**  $c \succ_{\mathcal{R}} a$ . By **Transfer**  $d \succeq_{\mathcal{R}} c$ . By **Speciesist Pareto**  $b \succ_{\mathcal{R}} d$ . By transitivity, we have a contradiction.  $\square$

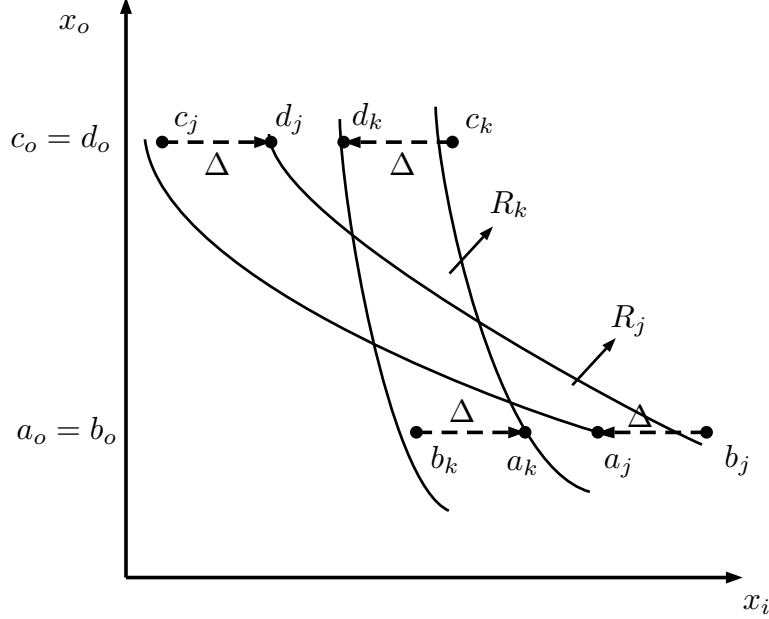


Figure 1: No SOF satisfies both **Speciesist Pareto** and **Transfer** on  $\mathcal{R}$ .

Proposition 1 shows that a consistent social evaluator cannot at the same time be speciesist and Paretian egalitarian. An evaluator who systematically respects unanimous preferences among humans (even when they harm animals) must give up egalitarianism among humans in the form of **Transfer**. Conversely, if an evaluator wants to uphold **Transfer**, she must resign herself to disagreeing with some unanimous preferences among humans.

As we show in the next section, an appropriate restriction of the directionality of Pareto improvements the evaluator is required to agree with provides a new way of resolving this conflict.<sup>5</sup> This is where our analysis differs from the analysis that could be done when instead of animal welfare, an ordinary public good is considered. For an

<sup>5</sup>To our knowledge, Sprumont (2012) and Van der Linden (2018) are the only other papers that consider weakening the Pareto principle as a way to overcome the incompatibility between egalitarianism and efficiency identified in Fleurbaey and Trannoy (2003).

ordinary public good, there is no reason to weaken the Pareto axiom, and therefore one has to seek a solution through weakening the egalitarian requirement. In contrast, here, the Pareto axiom looks strong because it does not take account of animals' interests. Incorporating these interests into the Pareto unanimity condition is the natural way to go if one wants to escape speciesism.

## 4 The possibility of a *non-speciesist* Paretian egalitarian

Consider now a *non-speciesist* social evaluator who displays concern for animal welfare alongside human preferences. For such an evaluator, a natural “first step” is to restrict the application of **Speciesist Pareto** to situations in which animal welfare does not decrease.

NON-SPECIESIST PARETO. For all  $\mathbf{R} \in \mathcal{D}$ , and all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$ , if  $\mathbf{x} R_i \mathbf{y}$  for all  $i \in H$  and  $x_o \geq y_o$ , then  $\mathbf{x} \succeq_{\mathbf{R}} \mathbf{y}$ ; if  $\mathbf{x} P_i \mathbf{y}$  for all  $i \in H$  and  $x_o \geq y_o$ , then  $\mathbf{x} \succ_{\mathbf{R}} \mathbf{y}$ .

Importantly, this is a *weakening* of **Speciesist Pareto** and it does not imply that a non-speciesist social evaluator only follows unanimous human preferences when animal welfare is not hurt. When animal welfare is hurt, this condition simply remains silent and does not require to consider the move to be a social welfare improvement, but it allows it.

Again, we stress that the reasons for **Non-speciesist Pareto** to be a natural restriction of **Speciesist Pareto** are specific to the context where  $x_o$  is interpreted as “animal welfare” and animal interests can be incorporated into the circle of relevant interests for which Paretian unanimity is checked.<sup>6</sup> **Non-speciesist Pareto** would not find such

---

<sup>6</sup>Note that the strict part of our axiom is not fully symmetric between animals and humans, as it requires a strict social preference when all humans have strict preferences even when animal welfare does not increase. This seems an acceptable relic of speciesism in our approach, and it also makes the formal analysis simpler.

a natural justification if  $x_o$  was another ordinary public good.

It is not hard to see that the restriction of **Speciesist Pareto** to **Non-speciesist Pareto** resolves the conflict with egalitarianism in the sense of **Transfer**. To see that **Speciesist Pareto** cannot be replaced by **Non-speciesist Pareto** in the the proof of Proposition 1, note that, in that proof, the latter axiom does not imply  $\mathbf{b} \succ_{\mathbf{R}} \mathbf{d}$ . **Transfer** and **Non-speciesist Pareto** are, for example, satisfied by the leximin SOF that is based on  $R^{AL}$  (i.e., animal-first lexicographic) reference preference, giving an absolute priority to animal welfare. Such a SOF prefers a change in allocation whenever animal welfare increases.

Less extreme and more “preference sensitive” SOFs also satisfy **Non-speciesist Pareto** and **Transfer**. Assuming that  $\mathcal{D} \subseteq \mathcal{R}^\theta$ , let  $\underline{MRS}(\mathbf{R})$  be the infimum of the marginal rates of substitution across all preferences in  $\mathbf{R}$  and all bundles. Formally,

$$\underline{MRS}(\mathbf{R}) := \inf\{m \in \mathbb{R}_+ \mid m = MRS(R_i; (x_o, x_i)) \text{ for some } (x_o, x_i) \in \mathbb{R}_+^2 \text{ and some } i \in H\}.$$

Also, for any  $\alpha \in \mathbb{R}_+$ , let  $R^\alpha$  be the linear preference with constant marginal rate of substitution  $\alpha$ . The leximin SOF based on the list of reference preferences  $(R^{\underline{MRS}(\mathbf{R})})_{\mathbf{R} \in \mathcal{R}^\theta}$  also satisfies **Non-speciesist Pareto** and **Transfer**. We call this SOF the **inf-MRS leximin** SOF.

Although it is less radical than the animal-leximin SOF, the inf-MRS leximin SOF remains very sensitive to extreme preferences. Indeed, the presence in the population of a single radically pro-animal individual can force the SOF to be arbitrarily close to an animal-leximin SOF. Because **Non-speciesist Pareto** seems like a weak axiom, and because it is unclear how egalitarianism among humans relates to social trade-offs between humans and animals, one may wonder whether more moderate SOFs can satisfy both **Non-speciesist Pareto** and **Transfer**. In other words, a natural question is whether the form of “dictatorship of deep ecologists” featured by the inf-MRS leximin SOF is a *necessary* consequence of combining **Non-speciesist Pareto** with transfer axioms.<sup>7</sup>

---

<sup>7</sup>Here and elsewhere, we use the word “dictatorship” in a sense that differs from its traditional definition in social choice theory. Consider the inf-MRS leximin SOF. It is not the case that the

Surprisingly, we find that an egalitarian social evaluator who embraces **Non-speciesist Pareto** *must* make the judgment of the most pro-animal human prevail. Specifically, we characterize the set of SOFs satisfying **Non-speciesist Pareto** together with a transfer and a mild continuity axiom, and show that they are based on a reference preference that is at least as pro-animal as the most pro-animal preference in the population.

To simplify the presentation, we first focus on a transfer axiom called **Equity** (see below) which is stronger than **Transfer**. In Section 6, we show that results similar to the ones we obtain in this section apply when **Equity** is replaced by **Transfer**. Also, as we show in Section 7, **Equity** follows from **Transfer** under the addition of natural robustness axioms, and replacing **Transfer** by **Equity** is in that sense innocuous.

Unlike **Transfer**, **Equity** requires that even “unbalanced” transfers from a human with high intrinsic welfare to a human with low intrinsic welfare be considered social improvements. This introduces a strong form of egalitarianism in the social ordering. Such an axiom is often called “Hammond Equity” (in reference to [Hammond, 1976](#)) in models featuring only human interests.

**EQUITY.** For all  $\mathbf{R} \in \mathcal{D}$ , and all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$ , if  $x_o = y_o$ , if there exist  $j, k \in H$  such that

$$y_j > x_j \geq x_k > y_k,$$

and if  $x_i = y_i$  for all  $i \neq j, k$ , then  $\mathbf{x} \succeq_{\mathbf{R}} \mathbf{y}$ .

We also adopt the following mild continuity requirement.<sup>8</sup>

---

human with the most pro-animal preference dictates her preference to the social evaluator in the traditional social choice sense. It can very well be that, while  $i$  has the most pro-animal preference and  $(x_o, x_i) R_i (y_o, y_i)$ , we have  $\mathbf{y} \succ_{\mathbf{R}} \mathbf{x}$  when  $\succeq$  is the inf-MRS leximin SOF. This occurs when, in spite of  $(x_o, x_i) R_i (y_o, y_i)$ , the worst-off human according to  $R_i$  is better-off according to  $R_i$  under  $\mathbf{y}$  than under  $\mathbf{x}$  (see Figure 2). By “dictatorship of the most pro-animal”, we mean that the most pro-animal human imposes the use of her preference as the reference preference through which the social evaluator must evaluate allocations.

<sup>8</sup>**Egalitarian Continuity** is a weak continuity requirement designed to be compatible with leximin SOFs. **Egalitarian Continuity** corresponds to what [Sprumont \(2012\)](#) calls “Weak Continuity” in his private good framework. We find the name “**Egalitarian Continuity**” to be more descriptive of the

EGALITARIAN CONTINUITY. For all  $\mathbf{R} \in \mathcal{D}$ , all  $x_o, y_o, w, z \in \mathbb{R}_+$ , and any sequence  $(z^n)_{n \in \mathbb{N}}$  with elements in  $\mathbb{R}_+$  converging to  $z$ , if

$$(y_o; z^k, \dots, z^k) \succeq_{\mathbf{R}} (x_o; w, \dots, w), \quad \text{for all } k \in \mathbb{N},$$

then  $(y_o; z, \dots, z) \succeq_{\mathbf{R}} (x_o; w, \dots, w)$ ; if

$$(y_o; z^k, \dots, z^k) \preceq_{\mathbf{R}} (x_o; w, \dots, w), \quad \text{for all } k \in \mathbb{N},$$

then  $(y_o; z, \dots, z) \preceq_{\mathbf{R}} (x_o; w, \dots, w)$ .

In Proposition 2 we show that, when combined, the weak concern for animals embodied in **Non-speciesist Pareto**, the egalitarianism among humans captured by **Equity**, and the mild **Egalitarian Continuity** imply a strong form of priority toward animals. Because human preferences remain the sole input of our SOFs and animals' preferences are not *directly* accounted for, this priority remains a function of human preferences (as in inf-MRS leximin SOFs). In particular, if no human cares for animals, the social evaluator is free to select a social ordering that predominantly favors human. However, as we show in Corollary 1, as soon as even a single human displays some “extreme” concern towards animals, this human’s extreme concerns prevail over the preferences of other humans who may not care for animals as much.

Before proving and discussing these two results let us note that, taken separately, each of the three axioms is fairly mild and normatively appealing. In particular, none of the three axioms is particularly biased in favor of animals. It is therefore all the more surprising that the *combination* of these three weak axioms implies giving such a strong priority to animal welfare. Note that removing either of the two main axioms allows to treat animals in a much less favorable way. For example, the *intrinsic welfare leximin SOF* that ranks allocations by applying the leximin criterion to the vector of human’s intrinsic welfare only satisfies **Equity** and **Egalitarian Continuity**, but violates **Non-speciesist Pareto**. This SOF allows to sacrifice arbitrarily large amounts of animal content of the axiom and adopt it throughout.

welfare in favor of arbitrarily small gains in humans' intrinsic welfare, regardless of humans' preferences.<sup>9</sup>

In order to state Proposition 2 formally, we introduce a class of SOF larger than the leximin class. For any allocation  $\mathbf{x}$ , let  $\min(\mathbf{x}) := \min_{i \in H} x_i$ . An SOF  $\succeq$  satisfies the **maximin property** if, for every profile  $\mathbf{R} \in \mathcal{D}$ , there exists a reference preference  $R(\mathbf{R})$  such that for all allocations  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$ ,

$$(x_o, \min(\mathbf{x})) P(\mathbf{R}) (y_o, \min(\mathbf{y})) \quad \text{implies} \quad \mathbf{x} \succ_{\mathbf{R}} \mathbf{y}. \quad (1)$$

If  $\succeq$  satisfies the maximin property and  $(R(\mathbf{R}))_{\mathbf{R} \in \mathcal{D}}$  is the list of reference preferences associated with  $\succeq$  as in (1), we say that  $\succeq$  is **based on**  $(R(\mathbf{R}))_{\mathbf{R} \in \mathcal{D}}$ .

Also, we say that preference  $R$  is at most as speciesist as preference  $R'$  if whenever a bundle is preferred according to  $R'$  to another bundle with at least as much animal welfare, the same is true according to  $R$ .<sup>10</sup> Formally,  $R$  is **at most as speciesist** as  $R'$  if for all  $(x_o, x_i), (y_o, y_i) \in \mathbb{R}_+^2$  such that  $y_o \geq x_o$ ,  $(y_o, y_i) R' (x_o, x_i)$  implies  $(y_o, y_i) R (x_o, x_i)$ .

**Proposition 2.** *On any domain  $\mathcal{D} \subseteq \mathcal{R}$ , if an SOF satisfies **Non-speciesist Pareto**, **Equity**, and **Egalitarian Continuity**, then*

---

<sup>9</sup>SOFs with a similar property can be constructed if **Equity** is dropped instead of **Non-speciesist Pareto**. For simplicity, consider the domain of linear preferences. Ascribe an arbitrary utility representation  $U^\alpha$  to any  $R^\alpha$  (recall that  $R^\alpha$  denotes the linear preference with constant marginal rate of substitution  $\alpha$ ). Let the preference of a generic human  $i \in H$  be denoted by  $R^{\alpha_i}$ . Also, let  $w(\alpha)$  be a strictly positive weighing function that is *increasing* in  $\alpha$ . Then consider the *weighted-utilitarian SOF* that ranks allocations as a function of  $\sum_{i \in H} w(\alpha_i) U^{\alpha_i}(x_o, x_i)$ . This SOF satisfies **Egalitarian Continuity** and **Non-speciesist Pareto** (because weights are strictly positive, the SOF actually satisfies the stronger **Speciesist Pareto**) but it violates **Equity**. The more  $w$  increases with  $\alpha$ , the more weighted-utilitarian SOFs give the priority to humans with a strong preference for intrinsic welfare. Therefore, for weighting functions  $w$  that are increasing enough, weighted-utilitarian SOFs can again sacrifice arbitrarily large amounts of animal welfare in favor of arbitrarily small gains in humans' intrinsic welfare, regardless of humans' preferences.

<sup>10</sup>"At most as speciesist" is equivalent to "at least as pro-animal".

(i) it satisfies the maximin property,<sup>11</sup> and

(ii) for every profile  $\mathbf{R} \in \mathcal{R}$ , the reference preference  $R(\mathbf{R})$  the SOF is based on is

(a) continuous in  $x_i$  and (b) at most as speciesist as  $R_i$  for all  $i \in H$ .

*Proof.* (i). The proof is by construction. Let  $\succeq$  be an SOF satisfying **Non-speciesist Pareto**, **Equity**, and **Egalitarian Continuity**. For any  $\mathbf{R} \in \mathcal{R}$ , define the binary relation  $R(\mathbf{R})$  on  $\mathbb{R}_+^2$  as follows: For all  $(x_o, x_i), (y_o, y_i) \in \mathbb{R}_+^2$ ,

$$(x_o, x_i) R(\mathbf{R}) (y_o, y_i) \quad \Leftrightarrow \quad (x_o; x_i, \dots, x_i) \succeq_{\mathbf{R}} (y_o; y_i, \dots, y_i). \quad (2)$$

Now, take any two allocations  $\mathbf{z}, \mathbf{w} \in \mathbb{R}_+^{h+1}$ . Suppose that

$$(z_o, \min(\mathbf{z})) P(\mathbf{R}) (w_o, \min(\mathbf{w})). \quad (3)$$

We need to show that  $\mathbf{z} \succ_{\mathbf{R}} \mathbf{w}$ .

By definition of  $R(\mathbf{R})$ , (3) implies

$$(z_o; \min(\mathbf{z}), \dots, \min(\mathbf{z})) \succ_{\mathbf{R}} (w_o; \min(\mathbf{w}), \dots, \min(\mathbf{w})).$$

By **Egalitarian Continuity**, there exists  $\delta > 0$  sufficiently small such that

$$(z_o; \min(\mathbf{z}), \dots, \min(\mathbf{z})) \succ_{\mathbf{R}} (w_o; \min(\mathbf{w}) + \delta, \dots, \min(\mathbf{w}) + \delta). \quad (4)$$

By **Non-speciesist Pareto**,

$$\mathbf{z} \succeq_{\mathbf{R}} (z_o; \min(\mathbf{z}), \dots, \min(\mathbf{z})). \quad (5)$$

By repeated application of **Equity**, there is  $\epsilon < \delta$  such that

$$(w_o; \min(\mathbf{w}) + \epsilon, \dots, \min(\mathbf{w}) + \epsilon) \succeq_{\mathbf{R}} \mathbf{w}. \quad (6)$$

By **Non-speciesist Pareto**,

$$(w_o; \min(\mathbf{w}) + \delta, \dots, \min(\mathbf{w}) + \delta) \succ_{\mathbf{R}} (w_o; \min(\mathbf{w}) + \epsilon, \dots, \min(\mathbf{w}) + \epsilon),$$

---

<sup>11</sup>Part (i) and its proof are inspired by Sprumont (2012, Proposition 1).

which combined with (4), (5) and (6), implies  $\mathbf{z} \succ_{\mathbf{R}} \mathbf{w}$ .

(ii). (a). Because  $\succeq$  satisfies **Egalitarian Continuity**, the reference preference  $R(\mathbf{R})$  is continuous in  $x_i$  by construction (see (2)).

(b). Suppose that for some  $i \in H$ ,  $R(\mathbf{R})$  is *not* at most as speciesist as  $R_i$ . That is, there are  $(x_o, x_i), (y_o, y_i) \in \mathbb{R}_+^2$  such that  $y_o \geq x_o$  and  $(y_o, y_i) R_i (x_o, x_i)$  but  $(x_o, x_i) P(\mathbf{R}) (y_o, y_i)$ . By (ii-a),  $R(\mathbf{R})$  is continuous in  $x_i$ . Therefore, there exists  $\epsilon$  sufficiently small such that  $(x_o, x_i) P(\mathbf{R}) (y_o, y_i + \epsilon)$ . Also,  $(y_o, y_i) R_i (x_o, x_i)$  implies  $(y_o, y_i + \epsilon) P_i (x_o, x_i)$  by monotonicity.

Let allocation  $\mathbf{z}$  be defined by  $z_o = x_o$ ,  $z_i = x_i$ , and  $z_j > x_i$  for all  $j \neq i$ . Let allocation  $\mathbf{w}$  be such that  $w_o = y_o$ ,  $w_i = y_i + \epsilon$ , and  $w_j > \max\{z_j, y_i + \epsilon\}$  for all  $j \neq i$ . Because  $w_o \geq z_o$  and  $w_j > z_j$  for all  $j \neq i$ , we have  $(w_o, w_j) P_j (z_o, z_j)$  for all  $j \neq i$ . Also,  $(w_o, w_i) = (y_o, y_i + \epsilon) P_i (x_o, x_i) = (z_o, z_i)$ . Thus, by **Non-speciesist Pareto**,  $\mathbf{w} \succ_{\mathbf{R}} \mathbf{z}$ .

Observe that  $i$  is the worst-off in both  $\mathbf{w}$  and  $\mathbf{z}$  according to  $R(\mathbf{R})$ . Therefore, because  $(x_o, x_i) P(\mathbf{R}) (y_o, y_i)$ , part (i) of the proof implies that  $\mathbf{z} \succ_{\mathbf{R}} \mathbf{w}$ , a contradiction.  $\square$

Naturally, any leximin SOF based on a (continuous) reference preference that is at most as speciesist as  $R_i$  for all  $i \in H$  satisfies the three axioms in Proposition 2. The following corollary shows how the three axioms imply a form of dictatorship from humans with extreme pro-animal preferences.

**Corollary 1.** *Suppose that some  $i \in H$  has animal-first lexicographic preference  $R_i = R^{AL}$ . Then the asymmetric part of any SOF obeying **Non-speciesist Pareto**, **Equity**, and **Egalitarian Continuity** satisfies: for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$ ,  $x_o > y_o$  implies  $\mathbf{x} \succ_{\mathbf{R}} \mathbf{y}$ .*

*Proof.* Corollary 1 is implied by Corollary 2 which is formally proven below.  $\square$

Corollary 1 considers the case of extreme pro-animal preferences. A similar result can be obtained with less extreme preferences: For every preference profile and for every *finite* set of allocations in which no two allocations have the same  $x_o$ , it is possible to



make the preference  $R_i$  of some human  $i$  sufficiently (but not totally) pro-animal, so that the allocations in the contemplated set are all ranked in terms of  $x_o$  only.

The results of this section are actually rather intuitive. If some members of the population have great willingness to pay for the public good (animal welfare), an improvement in animal welfare can rely on them paying most of the cost and produce a Pareto improvement among humans and animals included. But then these individuals may end up among the worst-off in the ranking of human (intrinsic) welfare, justifying to give them priority in redistribution. Therefore, the importance of their pro-animal preferences comes not from some bias in favor of animals in the social evaluation, but only from the combination of Paretianism, which accepts to let these people become poor for the sake of animals, and egalitarianism, giving priority to them in redistribution among humans only. The combination of such Pareto-improving enhancement of animal welfare and redistribution among humans ends up being equivalent to imposing equal contributions to the cause of animals on everyone in the population in order to please the most pro-animal members of the population.

## 5 Violations of **Speciesist Pareto**, and how to minimize them

Based on Proposition 2, it is not hard to show that SOFs that satisfy **Non-speciesist Pareto**, **Equity**, and **Egalitarian Continuity** violate **Speciesist Pareto**. We illustrated this point in Figure 2. Such violations of **Speciesist Pareto** should not come as a surprise given Proposition 1. When animals are only cared for as a function of pro-animal preferences among humans, violating **Speciesist Pareto** may, in fact, be desirable.

Yet, the violations of **Speciesist Pareto** imposed by the joint satisfaction of **Non-speciesist Pareto**, **Equity**, and **Egalitarian Continuity** can sometimes be troublesome. For example, configurations like the one depicted in Figure 2 can feature well-off humans who care more about animals than poorer humans, and whose preferences block

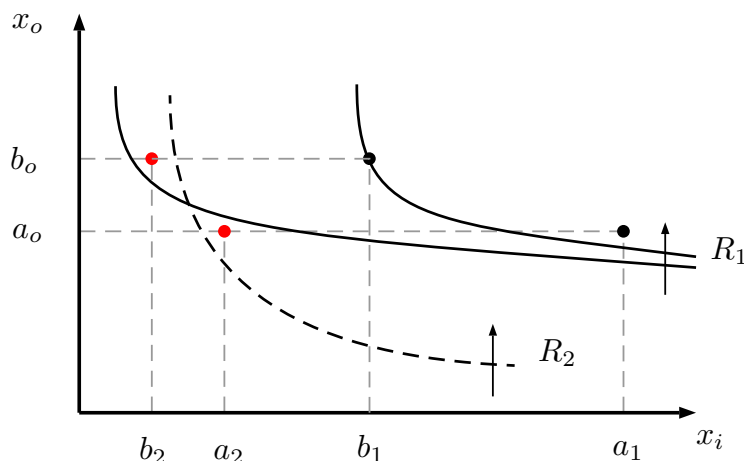


Figure 2: Suppose that an SOF satisfies **Non-speciesist Pareto**, **Equity**, and **Egalitarian Continuity** and that  $H = \{1, 2\}$ . Although  $\mathbf{a}$  is unanimously preferred to  $\mathbf{b}$ , the worst-off human according to  $R_1$  (human 2 in both allocations) is worse-off in  $\mathbf{a}$  than in  $\mathbf{b}$ . Because  $R_1$  is the least speciesist preference when comparing  $(a_o, a_2)$  and  $(b_o, b_2)$ , the SOF must therefore prefer  $\mathbf{b}$  to  $\mathbf{a}$  when the profile is  $\mathbf{R}$  (Proposition 2), contradicting **Speciesist Pareto**.

Pareto improvements that would have increased — at the expense of animal welfare — the intrinsic welfare of poorer humans. The SOFs characterized in Proposition 2 give an absolute priority to the worst-off for a given level of animal welfare, and respect the worst-off's preference when it is part of a unanimous agreement to increase animal welfare. But as the example in Figure 2 illustrates, these SOFs may impose a sacrifice on the worst-off for the sake of animal welfare, as valued by the affluents' preferences.

One must qualify this observation by noting that if the rich humans were in the situation of the worst-off, they would endorse the change and consider it an improvement for their own preference. Therefore, the above configuration occurs only if there is a correlation between preferences and affluence. The problem does not occur if every social group contains all sorts of preferences. In this case, an improvement in animal welfare at the expense of the worst-off would not be considered a social improvement if all the worst-off oppose it, as the least speciesist preference would also be present in

the worst-off group.

To be more precise, such an improvement in animal welfare at the expense of the worst-off would not *have to* be considered a social improvement. In principle, the SOFs characterized in Proposition 2 allow the social evaluator to select an SOF that goes *beyond* the least speciesist preference when it comes to protecting animals.<sup>12</sup> However, doing so implies further violations of **Speciesist Pareto**. Although violations of **Speciesist Pareto** are unavoidable, the above example shows that they can be problematic, and the evaluator may want to minimize them. That is, the evaluator may want to identify the SOF in the class characterized in Proposition 2 that is the least prone to violations of **Speciesist Pareto**.

Such an SOF has to rely on a reference preference that is as speciesist as possible among the preferences satisfying the condition in Proposition 2(ii). For any profile  $\mathbf{R} \in \mathcal{D}$ , let  $\mathcal{A}(\mathbf{R})$  denote the collection of preferences on  $\mathbb{R}_+^2$  that are at most as speciesist as  $R_i$  for all  $i \in H$ . In the Appendix, we show that, for every  $\mathbf{R} \in \mathcal{R}^c$ , there exists a **most speciesist** preference in every  $\mathcal{A}(\mathbf{R})$ , i.e., a preference  $R_{spec}(\mathbf{R}) \in \mathcal{A}(\mathbf{R})$  such that every preference in  $\mathcal{A}(\mathbf{R})$  is at most as speciesist as  $R_{spec}(\mathbf{R})$ .<sup>13</sup> Moreover, we show that  $R_{spec}(\mathbf{R})$  is convex if human preferences are convex.

In general, the reference preference  $R_{spec}(\mathbf{R})$  can be a complex function of  $\mathbf{R}$ . In some simple but important cases, however,  $R_{spec}(\mathbf{R})$  can be easily described. First, if there is one individual whose preference is at most as speciesist as others' preferences, then  $R_{spec}(\mathbf{R})$  coincides with this individual's preference. Second, at pairs  $(x_o, x_i)$  where every  $R_i$  is differentiable, the marginal rate of substitution of  $R_{spec}(\mathbf{R})$  corresponds to the lowest marginal rate of substitution *at that point* among the preferences in  $\mathbf{R}$ .

For any class  $\mathcal{S}$  of SOF, an element of the class  $\succeq^* \in \mathcal{S}$  is **most efficient in  $\mathcal{S}$**  if

---

<sup>12</sup>The characterization only puts a *lower*-bound on the evaluator's concern for animals, where the lower-bound is determined by the least speciesist preference in the population.

<sup>13</sup>Recall that  $\mathcal{R}^c \subset \mathcal{R}$  is the domain of all profiles the preferences of which are both strictly monotonic and continuous.

for all  $\mathbf{R} \in \mathcal{D}$ , for all  $\succeq \in \mathcal{S}$ , and for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$ , whenever  $(x_o, x_i) P_i (y_o, y_i)$  for all  $i \in H$  and  $\mathbf{x} \succ_{\mathbf{R}} \mathbf{y}$ , it is also the case that  $\mathbf{x} \succ_{\mathbf{R}}^* \mathbf{y}$ .

The next proposition highlights the leximin SOFs based on  $R_{spec}(\mathbf{R})$ , with respect to this efficiency criterion.

**Proposition 3.** *On domain  $\mathcal{R}^c$ , the leximin SOFs based on  $R_{spec}(\mathbf{R})$  are the most efficient SOFs in the class of leximin SOFs satisfying **Non-speciesist Pareto**, **Equity**, and **Egalitarian Continuity**.*

*Proof.* Consider any profile  $\mathbf{R} \in \mathcal{R}^c$  and any leximin SOF  $\succeq$  satisfying **Non-speciesist Pareto**, **Equity**, and **Egalitarian Continuity**. Suppose that  $\mathbf{x} P_i \mathbf{y}$  for all  $i \in H$  and  $\mathbf{x} \succ_{\mathbf{R}} \mathbf{y}$ . We prove that  $\mathbf{x} \succ_{R_{spec}(\mathbf{R})}^{lex} \mathbf{y}$ . If  $x_o \geq y_o$ , then  $\mathbf{x} \succ_{R_{spec}(\mathbf{R})}^{lex} \mathbf{y}$  follows directly from the fact that  $\succeq_{R_{spec}(\mathbf{R})}^{lex}$  satisfies **Non-speciesist Pareto**. Let us now focus on the case  $y_o > x_o$ .

Let  $R(\mathbf{R})$  be the reference preference  $\succeq$  is based on when the profile is  $\mathbf{R}$ . Note that by **Non-speciesist Pareto**, **Equity**, and **Egalitarian Continuity**,  $R(\mathbf{R})$  is at most as speciesist as  $R_{spec}(\mathbf{R})$  (Proposition 2). In order to derive a contradiction, suppose that  $\mathbf{y} \succeq_{R_{spec}(\mathbf{R})}^{lex} \mathbf{x}$ . To fix ideas, and without loss of generality, suppose that  $x_1 \leq \dots \leq x_h$  and  $y_1 \leq \dots \leq y_h$ .

By  $\mathbf{y} \succeq_{R_{spec}(\mathbf{R})}^{lex} \mathbf{x}$ , we have, either, for some  $k$ ,

$$\begin{aligned} (y_o, y_i) I_{spec}(\mathbf{R}) (x_o, x_i), & \quad \text{for all } i < k, \text{ and} \\ (y_o, y_k) P_{spec}(\mathbf{R}) (x_o, x_k). \end{aligned}$$

or

$$(y_o, y_i) R_{spec}(\mathbf{R}) (x_o, x_i), \quad \text{for all } i.$$

But because  $y_o > x_o$  and because  $R(\mathbf{R})$  is at most as speciesist as  $R_{spec}(\mathbf{R})$ , we then have either

$$\begin{aligned} (y_o, y_i) R(\mathbf{R}) (x_o, x_i), & \quad \text{for all } i < k, \text{ and} \\ (y_o, y_k) P(\mathbf{R}) (x_o, x_k). \end{aligned}$$

or

$$(y_o, y_i) R(\mathbf{R}) (x_o, x_i), \quad \text{for all } i.$$

This, in turn, implies  $\mathbf{y} \succeq_{\mathbf{R}} \mathbf{x}$ , contradicting  $\mathbf{x} \succ_{\mathbf{R}} \mathbf{y}$ . □

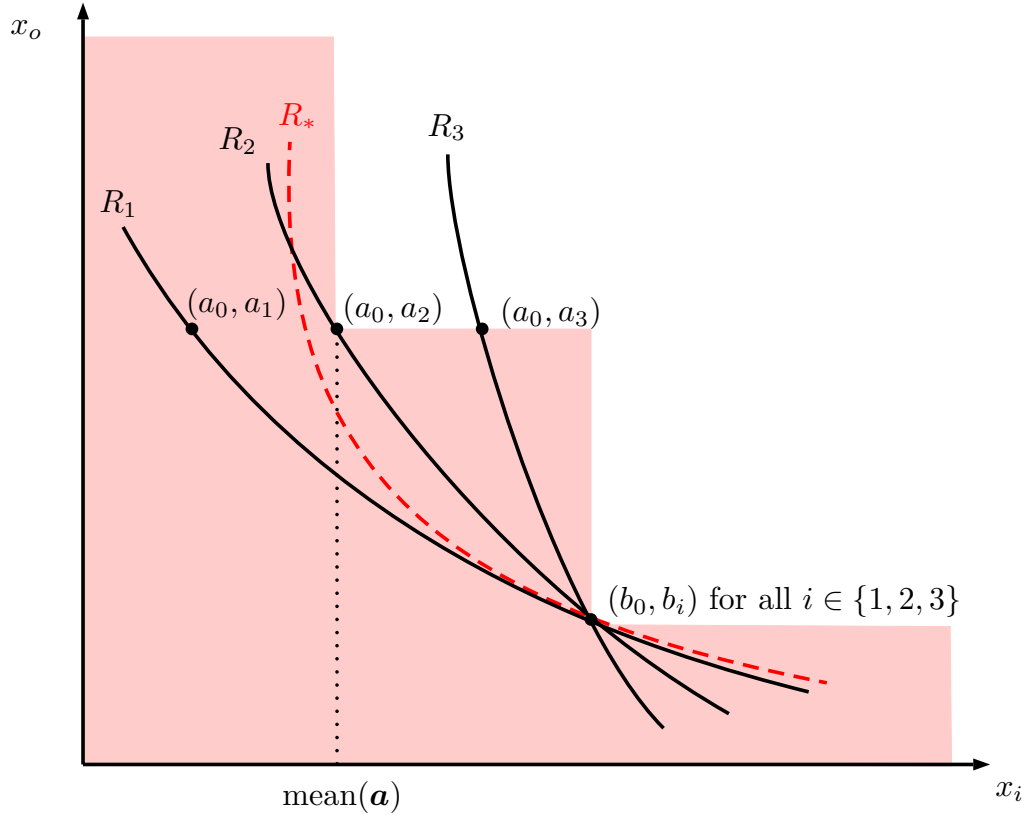


Figure 3:  $R_*$  is at most as speciesist on average as profile  $\mathbf{R} = (R_1, R_2, R_3)$  “for”  $\mathbf{a}$  and  $\mathbf{b}$ . For  $R_*$  to be at most as speciesist on average as profile  $\mathbf{R}$ , the property in the figure must be satisfied for all  $(b_o, b_i), (a_o, a_i) \in \mathbb{R}_+^{h+1}$  such that  $a_o \geq b_o$  and  $(a_o, a_i) R_i (b_o, b_i)$  for all  $i \in H$ . In general, if a preference is at most as speciesist on average as profile  $\mathbf{R} = (R_1, R_2, R_3)$  for  $\mathbf{a}$  and  $\mathbf{b}$ , its indifference curve passing through  $(b_o, b_i)$  must be contained in the red shaded area in the figure.

## 6 The dictatorship of extreme pro-animal preferences with Pigou-Dalton *transfers*

As we claimed above, our main conclusion — the dictatorship of extreme pro-animal preferences — does not depend on using the stronger **Equity** instead of **Transfer**. Naturally, **Transfer**, **Non-speciesist Pareto**, and **Egalitarian Continuity** jointly characterize a larger class of SOFs than when **Equity** is used in place of **Transfer**. As we show in Proposition 4, SOFs in this larger class also compare some allocations on the basis of a reference preference  $R(\mathbf{R})$ . This reference preference  $R(\mathbf{R})$  constraints the ranking of a smaller set of allocations than when **Equity** is used as an egalitarian axiom. However, these constraints are sufficient for Corollary 1 to generalize, which we show in Corollary 2.

For any allocation  $\mathbf{x}$ , let  $\text{mean}(\mathbf{x}) := (1/h) \sum_{i \in H} x_i$ . An SOF  $\succeq$  satisfies the **min-mean property** if for every profile  $\mathbf{R} \in \mathcal{D}$ , there exists a reference preference  $R(\mathbf{R})$  such that for all allocations  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$ ,

$$(x_o, \min(\mathbf{x})) P(\mathbf{R}) (y_o, \text{mean}(\mathbf{y})) \quad \text{implies} \quad \mathbf{x} \succ_{\mathbf{R}} \mathbf{y}.$$

Finally, let us say that preference  $R$  is **at most as speciesist on average** as profile  $\mathbf{R}$  if for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$  such that  $x_o \geq y_o$ , the unanimous preference  $(x_o, x_i) R_i (y_o, y_i)$  for all  $i \in H$  implies  $(x_o, \text{mean}(\mathbf{x})) R (y_o, \min(\mathbf{y}))$ , as illustrated in Figure 3.

**Proposition 4.** *On any domain  $\mathcal{D} \subseteq \mathcal{R}$ , if an SOF satisfies **Non-speciesist Pareto**, **Transfer**, and **Egalitarian Continuity**, then*

- (i) *it satisfies the min-mean property, and*
- (ii) *for every profile  $\mathbf{R} \in \mathcal{R}$ , the reference preference  $R(\mathbf{R})$  the SOF is based on is*
  - (a) *continuous in  $x_i$  and*
  - (b) *at most as speciesist on average as profile  $\mathbf{R}$ .*

*Proof.* (i). The proof is similar to that of Proposition 2. Let  $\succeq$  be an SOF satisfying **Non-speciesist Pareto**, **Transfer**, and **Egalitarian Continuity**. For any  $\mathbf{R} \in \mathcal{R}$ , define

the binary relation  $R(\mathbf{R})$  on  $\mathbb{R}_+^2$  as follows: For all  $(x_o, x_i), (y_o, y_i) \in \mathbb{R}_+^2$ ,

$$(x_o, x_i) R(\mathbf{R}) (y_o, y_i) \quad \Leftrightarrow \quad (x_o; x_i, \dots, x_i) \succeq_{\mathbf{R}} (y_o; y_i, \dots, y_i). \quad (7)$$

Now, take any two allocations  $\mathbf{w}, \mathbf{z} \in \mathbb{R}_+^{h+1}$ . Suppose that

$$(z_o, \min(\mathbf{z})) P(\mathbf{R}) (w_o, \text{mean}(\mathbf{w})). \quad (8)$$

We need to show that  $\mathbf{z} \succ_{\mathbf{R}} \mathbf{w}$ .

By definition of  $R(\mathbf{R})$ , (8) implies

$$(z_o; \min(\mathbf{z}), \dots, \min(\mathbf{z})) \succ_{\mathbf{R}} (w_o; \text{mean}(\mathbf{w}), \dots, \text{mean}(\mathbf{w})).$$

By **Egalitarian Continuity**, there exists  $\epsilon > 0$  sufficiently small such that

$$(z_o; \min(\mathbf{z}), \dots, \min(\mathbf{z})) \succ_{\mathbf{R}} (w_o; \text{mean}(\mathbf{w}) + \epsilon, \dots, \text{mean}(\mathbf{w}) + \epsilon). \quad (9)$$

By **Non-speciesist Pareto**,

$$\mathbf{z} \succeq_{\mathbf{R}} (z_o; \min(\mathbf{z}), \dots, \min(\mathbf{z})), \quad (10)$$

and

$$(w_o; \text{mean}(\mathbf{w}) + \epsilon, \dots, \text{mean}(\mathbf{w}) + \epsilon) \succ_{\mathbf{R}} (w_o; \text{mean}(\mathbf{w}), \dots, \text{mean}(\mathbf{w})). \quad (11)$$

Finally, observe that by repeated application of **Transfer**, we have

$$(w_o; \text{mean}(\mathbf{w}), \text{mean}(\mathbf{w}), \dots, \text{mean}(\mathbf{w})) \succeq_{\mathbf{R}} \mathbf{w},$$

which combined with (9), (10) and (11) again implies  $\mathbf{z} \succ_{\mathbf{R}} \mathbf{w}$ .

(ii). (a). Again, because  $\succeq$  satisfies **Egalitarian Continuity**, the reference preference  $R(\mathbf{R})$  is continuous in  $x_i$  by construction (see (7)).

(b). Suppose that  $R(\mathbf{R})$  is *not* at most as speciesist on average as profile  $\mathbf{R}$ . By definition, this means that there exists  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$  such that  $y_o \geq x_o$  and  $(y_o, y_i) R_i (x_o, x_i)$  for all  $i \in H$ , but  $(x_o, \min(\mathbf{x})) P(\mathbf{R}) (y_o, \text{mean}(\mathbf{y}))$ . By (ii-a).,  $R(\mathbf{R})$  is continuous in  $x_i$ . Therefore, there exists  $\epsilon > 0$  sufficiently small such that

defining  $\mathbf{e} := (\epsilon, \dots, \epsilon) \in \mathbb{R}^h$ , we have  $(x_o, \min(\mathbf{x})) P(\mathbf{R}) (y_o, \text{mean}(\mathbf{y} + \mathbf{e}))$ . By (i), this implies  $\mathbf{x} \succ \mathbf{y} + (0; \mathbf{e})$ .

Also, by monotonicity,  $(y_o, y_i) R_i (x_o, x_i)$  for all  $i \in H$  implies  $(y_o, y_i + \epsilon) P_i (x_o, x_i)$  for all  $i \in H$ . Then by **Non-speciesist Pareto**, we must have  $\mathbf{y} + (0; \mathbf{e}) \succ \mathbf{x}$ , a contradiction.

□

To our knowledge, Proposition 4 is the first result in the literature to identify the consequences of combining a transfer axiom “alone” with an efficiency requirement.<sup>14</sup> In Corollary 2, we show that even when equity is not implied, combining transfer and efficiency axioms can have far-reaching implications.

Naturally, any leximin SOF based on a (continuous) reference preference that is at most as speciesist as  $R_i$  for all  $i \in H$  satisfies the three axioms in Proposition 4.<sup>15</sup> However, some SOFs that do *not* rely on the leximin criterion also satisfy the three axioms in Proposition 4. For example, **Non-speciesist Pareto**, **Transfer**, and **Egalitarian Continuity** are satisfied by some **concave-utilitarian SOF**. An SOF  $\succeq^U$  is concave-utilitarian if for every profile  $\mathbf{R} \in \mathcal{D}$ , there exists a *concave* and continuous representation  $U^{R(\mathbf{R})}$  of a reference preference  $R(\mathbf{R})$  such that, for all allocations  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$ ,

$$\sum_{i \in H} U^{R(\mathbf{R})}(x_o, x_i) > \sum_{i \in H} U^{R(\mathbf{R})}(y_o, y_i) \quad \text{if and only if} \quad \mathbf{x} \succ_{\mathbf{R}}^U \mathbf{y}.$$

If the collection of reference preferences  $(R(\mathbf{R}))_{\mathbf{R} \in \mathcal{D}}$  a concave-utilitarian SOF  $\succeq^U$  based on is such that, for all  $\mathbf{R} \in \mathcal{D}$ ,  $R(\mathbf{R})$  is at most as speciesist as  $R_i$  for all  $i \in H$ ,

<sup>14</sup>Previous results either relied directly on an equity axiom (Sprumont, 2012), or justified relying on equity principles by showing that they are implied by transfer and efficiency axioms under additional robustness requirements (Fleurbaey and Maniquet, 2011). In contrast, Proposition 4 identifies implications of combining transfer and efficiency axioms when equity is *not* implied (as exemplified by the concave-utilitarian SOFs introduced below).

<sup>15</sup>Since these SOFs satisfy **Non-speciesist Pareto**, **Equity** and **Egalitarian Continuity**, and **Transfer** is implied by **Equity**.



then  $\succ^U$  satisfies **Non-speciesist Pareto**, **Transfer**, and **Egalitarian Continuity**. Observe that concave-utilitarian SOFs are not of the leximin kind and violate **Equity**.

The following result shows that, even for SOFs that violate **Equity**, combining transfer and efficiency axioms implies a form of dictatorship of humans with strong pro-animal preferences. Specifically, Corollary 2 shows that Corollary 1 does not depend on the use of **Equity** as an egalitarian axiom and also applies when **Transfer** is used instead (Corollary 2 therefore generalizes Corollary 1).

**Corollary 2.** *Suppose that some  $i \in H$  has a preference  $R_i = R^{AL}$ . Then the asymmetric part of an SOF  $\succeq$  obeying **Non-speciesist Pareto**, **Transfer**, and **Egalitarian Continuity** satisfies: for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$ ,  $x_o > y_o$  implies  $\mathbf{x} \succ_{\mathbf{R}} \mathbf{y}$ .*

*Proof.* The proof is illustrated in Figure 4. By part (ii) of Proposition 4 applied to small changes in  $x_o$ , we claim that the reference preference  $R(\mathbf{R})$  that  $\succeq$  is based on must care only about  $x_o$ . Without loss of generality, let human 1 have  $R_1 = R^{AL}$ .

Consider any  $a_o, b_o \in \mathbb{R}_+$  such that  $b_o > a_o$ . Observe that, by monotonicity, for any  $a_1 \in \mathbb{R}_+$ , we have  $(b_o, a_1) R_i (a_o, a_1)$  for all  $i \in H$ . Also, because  $R_1 = R^{AL}$ ,  $(b_o, 0) P_1 (a_o, a_1)$ .

Now, suppose that  $a_1 > 0$  and let  $\mathbf{w} := (b_o; 0, a_1, \dots, a_1)$  and  $\mathbf{z} := (a_o; a_1, \dots, a_1)$ . We then have  $\mathbf{w} R_i \mathbf{z}$  for all  $i \in H$ . By Proposition 4, this implies  $(w_o, \text{mean}(\mathbf{w})) R(\mathbf{R}) (z_o, \min(\mathbf{z}))$  which can be re-written as  $(b_o, [(h-1)/h]a_1) R(\mathbf{R}) (a_o, a_1)$ .

The inequality  $[(h-1)h]a_1 < a_1$  is independent of  $b_o$ . Therefore, as  $b_o \rightarrow a_o$ , the indifference curve<sup>16</sup> of  $R(\mathbf{R})$  becomes flatter and flatter on the (positive length) horizontal stretch between bundles  $(a_o, [(h-1)/h]a_1)$  and  $(a_o, a_1)$ . In the limit, the indifference curve of  $R(\mathbf{R})$  must be flat between  $(a_o, [(h-1)/h]a_1)$  and  $(a_o, a_1)$ . Because this is true for any  $(a_o, a_1) \in \mathbb{R}_+^2$ , the indifference curves of  $R(\mathbf{R})$  must be flat everywhere, which proves the claim.

<sup>16</sup>This is an abuse of language, since  $R(\mathbf{R})$  need not be continuous in  $(x_o, x_i)$ , even if it is continuous in  $x_i$ . By “indifference curve” here, we mean the lower boundary of the upper contour set.

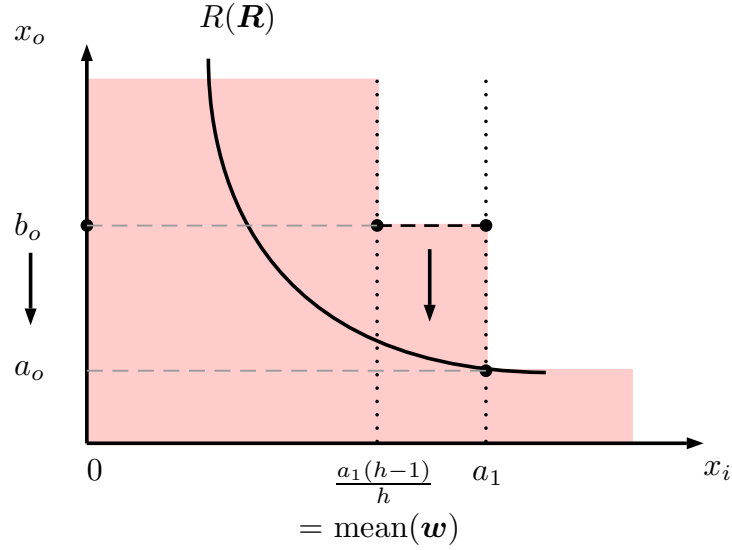


Figure 4: Proof of Corollary 2. The indifference curve passing through  $(a_o, a_1)$  cannot lie outside of the red area. This implies that, as  $b_o \rightarrow a_o$ , the indifference curve must be flat between bundles  $(a_o, [(h-1)/h]a_1)$  and  $(a_o, a_1)$ .

Finally, because  $R(\mathbf{R})$  has flat indifference curves, it must be that for any  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{h+1}$ ,  $x_o > y_o$  implies  $(x_o, \min(\mathbf{x})) P(\mathbf{R}) (y_o, \text{mean}(\mathbf{y}))$ , which in turn implies that  $\mathbf{x} \succ_{\mathbf{R}} \mathbf{y}$ .  $\square$

As in Corollary 1, Corollary 2 considers the case of extreme pro-animal preferences. Again, a similar result can be obtained with less extreme preferences: for every profile of preferences and for every *finite* set of allocations in which no two allocations have the same  $x_o$ , it is possible to make the preference  $R_i$  of some human  $i$  sufficiently (but not totally) pro-animal, so that the allocations in the contemplated set are all ranked in terms of  $x_o$  only.

## 7 From Transfer to Dominance Aversion

In this final section we show how, following standard techniques, **Equity** can be derived from **Transfer** when **Non-speciesist Pareto** is satisfied and additional robustness condi-

tions are imposed. In particular, the results in this section imply that the characterization in Proposition 2 holds when **Equity** is replaced by **Transfer** and the appropriate robustness axioms are added (a slight strengthening of the continuity axiom is also required, see below).

The first robustness axiom requires defining the notion of (closed) upper contour set, i.e., the set of bundles that are weakly preferred to the bundle under consideration:

$$C^U((x_o, x_i), R_i) := \{(z_o, z_i) \in \mathbb{R}_+^2 \mid (z_o, z_i) R_i (x_o, x_i)\}.$$

The axiom says that the only relevant information for the social assessment of two allocations is the upper contour sets at these allocations. The motivation is that, when comparing two allocations, it may be important to know about the bundles that an individual prefers to her bundles in these allocations. But detailed information about an individual's preference over pairs of bundles that do not include her bundle in either of the allocations should be irrelevant. It is worth noting that this informational restriction is satisfied by the most common approaches to constructing social preferences, including those used in cost-benefit analysis and involving compensating variations and equivalent variations.

**UNCHANGED CONTOUR INDEPENDENCE** (HANSSON, 1973). For all  $\mathbf{R}, \mathbf{R}' \in \mathcal{D}$ , and all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$ , if  $C^U((x_o, x_i), R_i) = C^U((x_o, x_i), R'_i)$  and  $C^U((y_o, y_i), R_i) = C^U((y_o, y_i), R'_i)$  for all  $i \in H$ , then

$$\mathbf{x} \succeq_{\mathbf{R}} \mathbf{y} \quad \text{if and only if} \quad \mathbf{x} \succeq_{\mathbf{R}'} \mathbf{y}.$$

We also slightly strengthen the continuity axiom. A sequence of allocations  $(\mathbf{x}^n)_{n \in \mathbb{N}}$  is **order-preserving** if the ordering of intrinsic welfare among humans whom intrinsic welfare changes somewhere in the sequence is preserved throughout the sequence. Formally, sequence  $(\mathbf{x}^n)_{n \in \mathbb{N}}$  is order-preserving if for all  $j, k \in H$  for whom there exists  $m, p, q, r \in \mathbb{N}$  such that  $x_j^m \neq x_j^p$  and  $x_k^q \neq x_k^r$ , it is the case that either  $x_j^n \geq x_k^n$  for all  $n \in \mathbb{N}$ , or  $x_j^n \leq x_k^n$  for all  $n \in \mathbb{N}$ .

ORDER-PRESERVING CONTINUITY. For all  $\mathbf{R} \in \mathcal{D}$ , all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$ , and any *order preserving* sequence of allocations  $(\mathbf{x}^n)_{n \in \mathbb{N}}$  with elements in  $\mathbb{R}_+^{h+1}$  converging to  $\mathbf{x}$ , if  $\mathbf{x}^k \succeq_{\mathbf{R}} \mathbf{y}$ , for all  $k \in \mathbb{N}$ , then  $\mathbf{x} \succeq_{\mathbf{R}} \mathbf{y}$ ; if  $\mathbf{x}^k \preceq_{\mathbf{R}} \mathbf{y}$ , for all  $k \in \mathbb{N}$ , then  $\mathbf{x} \preceq_{\mathbf{R}} \mathbf{y}$ .

The next proposition shows that **Non-speciesist Pareto**, **Transfer**, **Order-Preserving Continuity**, and **Unchanged Contour Independence** jointly imply **Equity**.

**Proposition 5.** *On domain  $\mathcal{R}$ , if an SOF satisfies **Non-speciesist Pareto**, **Transfer**, **Order-Preserving Continuity**, and **Unchanged Contour Independence**, then it satisfies **Equity**.*

*Proof. Step 1:* If  $\mathbf{a}, \mathbf{b} \in \mathbb{R}_+^{h+1}$  are such that (i)  $b_o > a_o$ , (ii)  $a_j > b_j > b_k > a_k$ , and (iii)  $b_i = a_i$  for all  $i \neq j, k$ , then  $\mathbf{b} \succeq_{\mathbf{R}} \mathbf{a}$ .<sup>17</sup>

Figure 5 illustrates this step. The figure represents changes in the bundles of  $j$  and  $k$  only, and it is assumed that  $x_i$  stays put for  $i \neq j, k$ , implying that the satisfaction of humans other than  $j$  and  $k$  does not decrease when  $x_o$  rises. We let  $\mathbf{c}, \dots, \mathbf{e}$  denote the allocations constructed in this way. In particular, as the figure illustrates,  $d_k - c_k = c_j - d_j$  and  $b_k - e_k = e_j - b_j$ .

In order to derive a contradiction, assume that  $\mathbf{a} \succ_{\mathbf{R}} \mathbf{b}$ .

By **Unchanged Contour Independence**, one can change the indifference curve for  $j$  at  $c_j$ , as the dotted curve in the figure, while keeping indifference curves at  $\mathbf{a}$  and  $\mathbf{b}$  unchanged, thus generating a new profile  $\mathbf{R}'$  for which we still have  $\mathbf{a} \succ_{\mathbf{R}'} \mathbf{b}$ .

By **Non-speciesist Pareto**, moving from  $\mathbf{a}$  to  $\mathbf{c}$  is a strict improvement for  $\succeq_{\mathbf{R}'}$ . One therefore has  $\mathbf{c} \succ_{\mathbf{R}'} \mathbf{b}$  (by transitivity).

By **Unchanged Contour Independence**, one can change the indifference curve for  $j$  at  $e_j$ , as the dashed curve in the figure, while keeping indifference curves at  $\mathbf{c}$  and  $\mathbf{b}$  unchanged, thus generating a new profile  $\mathbf{R}''$  for which we still have  $\mathbf{c} \succ_{\mathbf{R}''} \mathbf{b}$ .

By **Transfer**, moving from  $\mathbf{c}$  to  $\mathbf{d}$  is a weak improvement for  $\succeq_{\mathbf{R}''}$ , so that  $\mathbf{d} \succ_{\mathbf{R}''} \mathbf{b}$ .

By **Non-speciesist Pareto**, moving from  $\mathbf{d}$  to  $\mathbf{e}$  is a strict improvement for  $\succeq_{\mathbf{R}''}$ , so that  $\mathbf{e} \succ_{\mathbf{R}''} \mathbf{b}$ .

---

<sup>17</sup>This differs from the requirement of **Equity** because of (i).

By **Transfer**, moving from  $\mathbf{e}$  to  $\mathbf{b}$  is a weak improvement for  $\succeq_{\mathbf{R}'}$ , so that  $\mathbf{b} \succ_{\mathbf{R}'} \mathbf{a}$ .

But by **Unchanged Contour Independence**, since the indifference curves for  $j$  and  $k$  at  $\mathbf{a}$  and  $\mathbf{b}$  are the same under both  $\mathbf{R}$  and  $\mathbf{R}'$ , this implies  $\mathbf{b} \succ_{\mathbf{R}} \mathbf{a}$ , which contradicts  $\mathbf{a} \succ_{\mathbf{R}} \mathbf{b}$ .

**Step 2:** Step 1 holds even if (i) is replaced by (i')  $b_o = a_o$ .

Consider any allocations  $\mathbf{a}, \mathbf{b}$  such that (i'), (ii), and (iii) hold. Construct a sequence of allocations  $(\mathbf{b}^n)_{n \in \mathbb{N}}$  such that for all  $n \in \mathbb{N}$ , (a)  $b_i^n = b_i$  for all  $i \in H$ , and (b)  $b_o^n > a_o$  with  $b_o^n \rightarrow a_o$  as  $n$  tends to infinity. By construction, for every  $n \in \mathbb{N}$ ,  $b_o^n > a_o$ , and  $a_j > b_j^n > b_k^n > a_k$ , whereas  $b_i^n = a_i$  for all  $i \neq j, k$ . Thus, by Step 1, for every element  $\mathbf{b}^n$  of the sequence, one has  $\mathbf{b}^n \succeq_{\mathbf{R}} \mathbf{a}$ . By **Order-Preserving Continuity**, one therefore also has  $\mathbf{b} \succeq_{\mathbf{R}} \mathbf{a}$ , which proves **Equity**.  $\square$

**Equity** can also be derived using another robustness axiom. The following separability condition says that an individual who has the same bundle in two allocations and is therefore unconcerned by the change from one allocation to the other does not count, i.e., his situation (bundle and preference) could change without affecting the social preference. The axiom is weak in the sense that it only requires separability between allocations with the same level of animal welfare.

**SEPARABILITY.** For all  $\mathbf{R} \in \mathcal{D}$  and all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{h+1}$ , if  $x_o = y_o$  and there exists  $i \in H$  such that  $x_i = y_i$ , then for all  $R'_i \in \mathcal{D}_i$  and all  $w_i \in \mathbb{R}_+$  we have

$$\mathbf{x} \succeq_{\mathbf{R}} \mathbf{y} \quad \text{if and only if} \quad (x_o; w_i, x_{-i}) \succeq_{(R'_i, R_{-i})} (y_o; w_i, y_{-i}).$$

**Proposition 6.** *On domain  $\mathcal{R}$ , if  $h \geq 3$  and an SOF satisfies **Non-speciesist Pareto**, **Transfer**, **Order-Preserving Continuity**, and **Separability**, then it satisfies **Equity**.*

*Proof.* Consider any  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{h+1}$  and let  $\rho := b_j - b_k$ ,  $\Delta_j := a_j - c_j$ , and  $\Delta_k := b_k - d_k$ .

**Case 1.**  $\rho > \max\{\Delta_j, \Delta_k\}$ .

**Step 1.** If  $\mathbf{a}, \mathbf{b} \in \mathbb{R}_+^{h+1}$  are such that (i)  $a_o > b_o$ , (ii)  $a_j > b_j > a_g > b_g > b_k > a_k$ , and (iii)  $b_i = a_i$  for all  $i \neq g, j, k$ , then for some preference  $R'_g$ , we have  $\mathbf{b} \succeq_{(R'_g, R_{-g})} \mathbf{a}$ .<sup>18</sup>

<sup>18</sup>This is different from **Equity** because of (i), (ii) and because, potentially,  $(R'_g, R_{-g}) \neq \mathbf{R}$ .

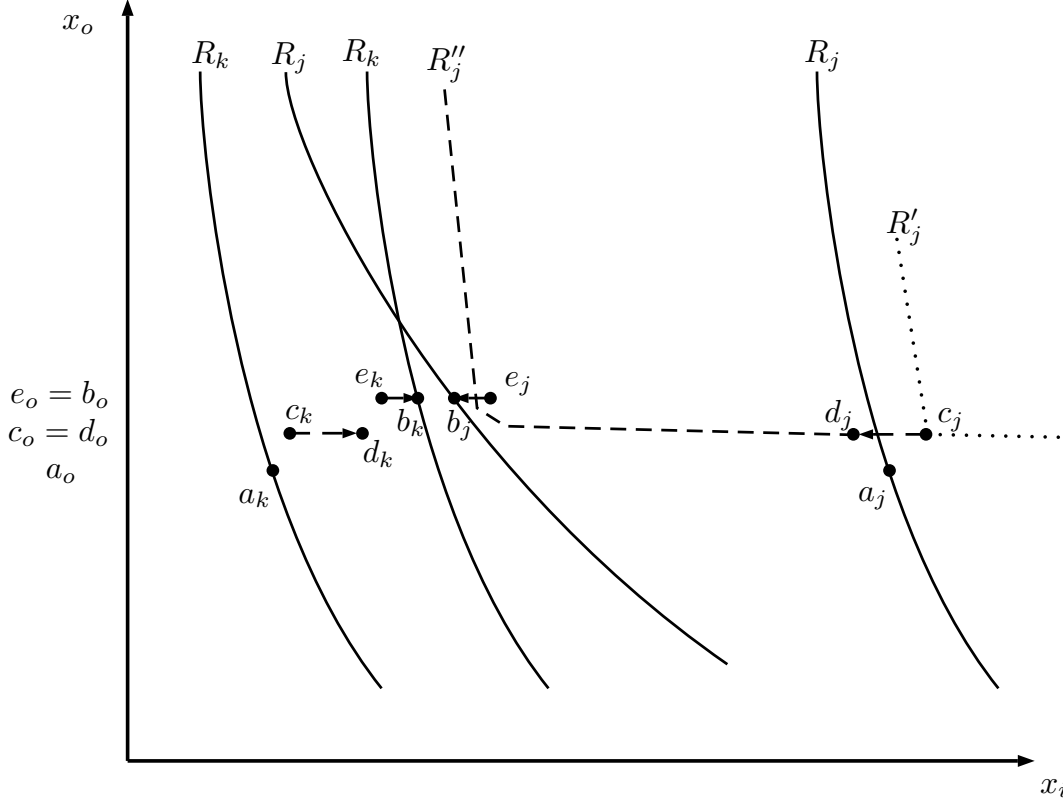


Figure 5: **Non-speciesist Pareto, Transfer, Egalitarian Continuity, Unchanged Contour Independence** imply **Equity**.

The proof in this case is illustrated in Figure 6. The figure represents the particular preference  $R'_g$  as well as changes in the bundles of  $g$ ,  $j$ , and  $k$  only. It is assumed that  $x_i$  stays put for all  $i \neq g, j, k$ , implying that the satisfaction of humans other than  $g$ ,  $j$  and  $k$  does not decrease when  $x_o$  rises. We let  $\mathbf{c}$  and  $\mathbf{d}$  denote the allocations constructed in this way. In particular, as the figure illustrates,  $a_j - c_j = c_g - a_g = \Delta_j$  and  $b_k - d_k = d_g - b_g = \Delta_k$ . Such a construction is made possible by Case 1's assumption that  $\rho > \max\{\Delta_j, \Delta_k\}$ .

By **Transfer**,  $\mathbf{c} \succeq_{(R'_g, R_{-g})} \mathbf{a}$ . By **Non-speciesist Pareto**,  $\mathbf{d} \succ_{(R'_g, R_{-g})} \mathbf{c}$ . Finally, by **Transfer** again,  $\mathbf{b} \succeq_{(R'_g, R_{-g})} \mathbf{d}$ , which by transitivity implies  $\mathbf{b} \succeq_{(R'_g, R_{-g})} \mathbf{a}$ .

**Step 2.** If  $\mathbf{a}, \mathbf{b} \in \mathbb{R}_+^{h+1}$  are such that (i')  $a_o = b_o$ , (ii')  $a_j > b_j > a_g = b_g > b_k > a_k$ ,

and (iii)  $b_i = a_i$  for all  $i \neq g, j, k$ , then for some preference  $R'_g$ , we have  $\mathbf{b} \succeq_{(R'_g, R_{-g})} \mathbf{a}$ .<sup>19</sup>

Construct a sequence of allocations  $(\mathbf{b}^n)_{n \in \mathbb{N}}$  such that for all  $n \in \mathbb{N}$ , (i), (ii), and (iii) are satisfied, and such that (a)  $b_o^n$  converges to  $a_o$ , and (b)  $b_g^n$  converges to  $a_g$ . By Step 1, for every element  $\mathbf{b}^n$  of the sequence, one has  $\mathbf{b}^n \succeq_{(R'_g, R_{-h})} \mathbf{a}$ . By **Order-Preserving Continuity**, one therefore also has  $\mathbf{b} \succeq_{(R'_g, R_{-g})} \mathbf{a}$ .

**Step 3.** Finally, by **Separability** and because Step 2 holds for arbitrary  $\mathbf{a}$  and  $\mathbf{b}$  satisfying (i'), (ii'), and (iii), if (i'')  $b_o = a_o$ , (ii'')  $b_j > a_j > a_k > b_k$ , and (iii')  $b_i = a_i$  for all  $i \neq j, k$ , then  $\mathbf{a} \succeq_{\mathbf{R}} \mathbf{b}$ , which proves **Equity**.

**Case 2.**  $\rho \leq \max\{\Delta_j, \Delta_k\}$ .

This second case follows by repeated application of Case 1. Simply consider any  $m \in \mathbb{N}$  such that  $\rho > (1/m) \max\{\Delta_j, \Delta_k\}$ . By applying Case 1  $m$  times to redistributions where  $\Delta_k/m$  is taken away from  $k$  and  $\Delta_j/m$  is given to  $j$ , one again obtains  $\mathbf{a} \succeq_{\mathbf{R}} \mathbf{b}$ , which proves **Equity**.  $\square$

## 8 Conclusion

In this paper, it has been shown that if the Pareto principle is satisfied whenever animals are not hurt and a certain degree of egalitarianism among humans is adopted, then the least speciesist preference in the population has a strong sway over social preferences. This result offers a surprising connection between egalitarianism among humans and pro-animal preferences. The result is actually not so counter-intuitive: Humans with the least speciesist preferences are the most likely to end up among the worst-off when their greater willingness to pay for animal welfare leads them to pay more than others for improvements in the fate of other species. Being among the worst-off, these individuals then become a social priority, and their preferences therefore become important as well. We have shown that this “dictatorship of the least speciesist” extends even to the case in which the degree of egalitarianism is mild.

---

<sup>19</sup>This differs from Step 1 because  $b_o = a_o$  in (i'), and because  $a_g = b_g$  in (ii').

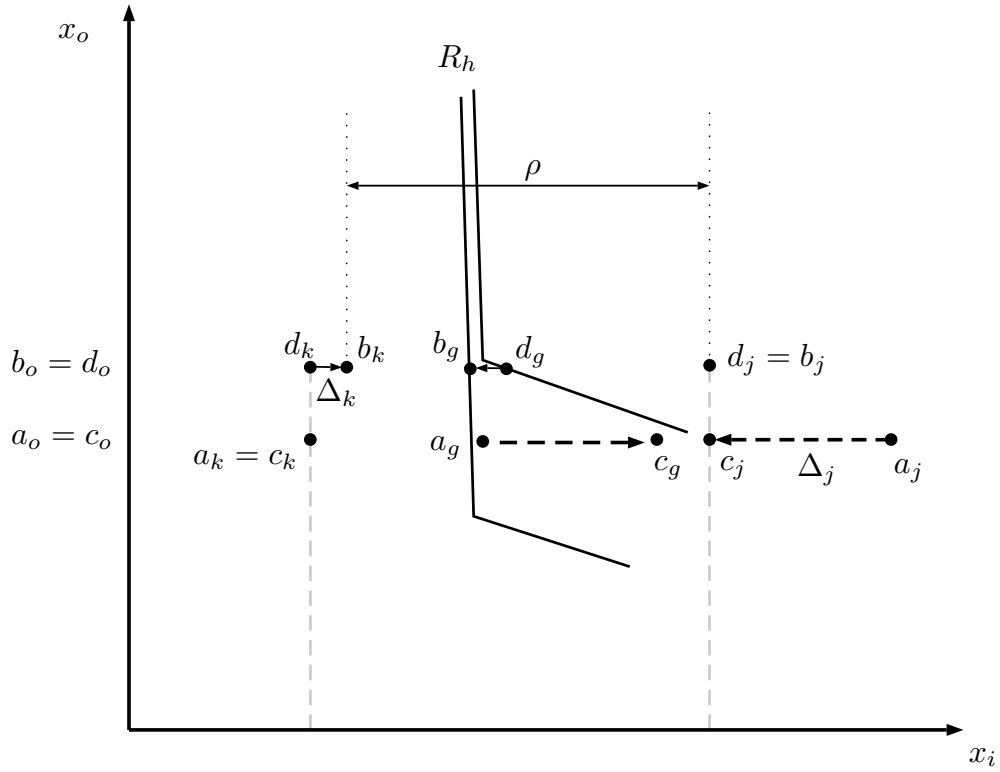


Figure 6: **Non-speciesist Pareto, Transfer, Egalitarian Continuity, Separability** imply **Equity** when  $\rho > \max\{\Delta_j, \Delta_k\}$ .

In particular, giving an absolute social priority to the worst-off is *not* required for highly pro-animal preferences to have a determinant impact on social preferences.

As mentioned in the beginning, an important limitation of this paper lies with the fact that animal welfare has been treated as a unidimensional index, as if all humans in the population agreed on how to assess the situation of other species. However, the intuition provided in the previous paragraph shows that the bulk of the analysis remains valid even when this assumption is relaxed, in spite of the complications introduced by multidimensionality of animal welfare. Indeed, if there is one species some humans care a lot about, these humans will again become a social priority when paying to improve the welfare of their beloved species. A strong preference in the population in favor of *any* species is therefore enough to carry the social ordering in the direction



of this species' welfare. If some species are less liked than others, however, the social ordering may give them less weight in the evaluation of allocations.

The conclusions of this paper rely on the combination of an egalitarian axiom (**Transfer** or **Equity**) which focuses on consumption inequalities among humans and a Pareto axiom (**Non-speciesist Pareto**) which encompasses animals in its scope. One can worry that by requiring that consumption inequalities be alleviated at *all* levels of animal welfare and regardless of preferences, our egalitarian axioms implicitly favor pro-animal humans in the population. In particular, the egalitarian axioms we study in this paper may recommend transfers from richer speciesist humans towards poorer pro-animal humans *even when* animal welfare is high. These transfers may be troublesome as the pro-animal humans who benefit from these transfers could be viewed as better-off than their richer speciesist counterparts by virtue of their stronger preferences for animals and the fact that animal welfare is already high.

Alternative SOFs exist that take these concerns into account by allowing pro-animal humans to be considered better-off than the more speciesist humans even if they consume less, provided animal welfare is sufficiently large. One example is the SOF that fixes a reference level of animal well-being and applies the leximin ordering to humans' "equivalent consumption" at that reference level of animal well-being (defined as the level of consumption which, if combined with the reference level of animal well-being, would leave the human indifferent with her current bundle). Over the set of allocations for which the equivalent consumption is defined for every individual, this SOF is easily characterized by the combination of **Speciesist Pareto** and a weakened **Equity** axiom that applies only when animal welfare is at the reference level.

When animal welfare is above the reference level, this SOF gives priority to the *least* pro-animal preferences and prevents transfers like the ones described in the previous paragraph which some may find troublesome. Note, however, that this comes at the cost of the "equivalent consumption" SOF recommending equally troublesome *regressive* transfers in favor of pro-animal humans. For example, for levels of animal welfare

that are below the reference level, the “equivalent consumption” SOF may recommend that consumption be transferred *away* from poorer speciesist humans in favor of richer pro-animal humans (as a form of “compensation” for the fact that pro-animal humans “suffer more” from the low level of animal welfare than the speciesist).

Perhaps the strongest limitation of this paper is that it does not grant any intrinsic value to other species. Their priority in social preferences is completely conditional on human willingness to promote their welfare (although we show that the pro-animal preference of *even a single human* can force the social evaluator to care for other species). Analyzing non-human animals as members of the population with their own preferences (if not values) appears a challenging task, but has already found advocates (e.g., [Kymlicka and Donaldson, 2011](#)). We leave this intriguing question for future research.

## Appendix

**Existence of a most speciesist preference among the preferences that are at most as speciesist as  $R_i$  for all  $i \in H$ .**

**Proposition 7.** *For any given profile  $\mathbf{R} \in \mathcal{R}$ , let  $\mathcal{A}(R)$  denote the collection of preferences over  $\mathbb{R}_+^2$  that are at most as speciesist as  $R_i$  for all  $i \in H$ . This collection has a well-defined most speciesist member, which is convex if every  $R_i$  is convex.*

*Proof.* If there is  $i \in H$  such that  $R_i = R^{AL}$ , this individual has preferences that are at most as speciesist as everyone else, and then  $R^{AL}$  itself is the most speciesist among the preferences that are at most as speciesist as every  $R_i$ . In the rest of the proof, we restrict attention to the case in which all individual preferences are continuous and monotonic.

The first part of the proof consists in showing that the relation “at most as speciesist as”, which we will denote by  $\triangleright$ , generates a lattice structure on the set of continuous

and strictly monotonic preferences. The proposition states that every collection of  $h$  preferences has a meet in the domain. This is indeed true if we have a lattice.

It suffices to show that every pair of preferences has a meet and a join for  $\supseteq$ . We focus on the join, since the proof for the meet is essentially identical.

Continuous and strictly monotonic preferences have strictly decreasing indifference curves, which are therefore differentiable a.e. (almost everywhere). Let  $R, R'$  be two preferences and let  $I(x, R)$  denote the indifference curve of  $R$  that contains  $x = (x_o, x_i) \in \mathbb{R}_+^2$ . Let  $s(x, R)$  denote the derivative of the graph of the  $I(x, R)$  (i.e., the slope of the curve, which must be negative a.e. by strict monotonicity of preferences).

The construction of the indifference map of the join  $R \vee R'$  goes as follows:

1. At a point  $x$  where  $s(x, R) > s(x, R')$ , let the curve of  $R \vee R'$  follow  $I(x, R)$  for  $y_i > x_i$  until the condition  $s(y, R) > s(y, R')$  is no longer satisfied.
2. Symmetrically, when  $s(x, R) < s(x, R')$ , let the curve follow  $I(x, R')$  for  $y_i > x_i$  until the condition  $s(y, R) > s(y, R')$  is no longer satisfied.
3. At a point  $x$  where  $s(x, R) = s(x, R')$ , let the curve follow  $I(x, R)$  for  $y_i > x_i$  until the condition  $s(x, R) = s(x, R')$  is no longer satisfied.

This construction defines the indifference map of  $R \vee R'$  a.e., and by continuity the construction is extended to the whole space.

This construction is unique because it is unique in the vicinity of a.e. point, and every point that is not included in steps 1-3 has a single curve that tends toward it (indeed, it has only one curve for each preference  $R, R'$  that tends toward it), so that the continuity extension is also unique.

A curve of this map is continuous because it is made of continuous bits of indifference curves and the continuity extension imposes continuity at the points where steps 1-3 do not apply.

Curves of  $R \vee R'$  cannot cross because this would mean that at a particular point  $x$ , the construction is not unique.

A curve of  $R \vee R'$  is decreasing a.e. by steps 1-3 because  $s(x, R) < 0$  for all  $x$  and all  $R$  in the domain. It is therefore decreasing after the continuity extension.

In conclusion,  $R \vee R'$  is in the domain of continuous, strictly monotonic preferences.

It is straightforward to check that  $R'' \supseteq R \vee R'$  for all  $R''$  such that  $R'' \supseteq R$  and  $R'' \supseteq R'$ . Indeed, the fact that  $R'' \supseteq R$  and  $R'' \supseteq R'$  implies that  $s(x, R'') \geq \max\{s(x, R), s(x, R')\}$  for almost every  $x$  (i.e., every  $x$  where this is defined), and it is never the case that  $s(x, R \vee R') > \max\{s(x, R), s(x, R')\}$ .

This concludes the proof of existence of a join.

When both  $R$  and  $R'$  are convex, their indifference curves have non-decreasing slopes, and therefore this holds for the curves of  $R \vee R'$  too, which are made of parts of these curves. When the join is taken for a finite set of more than two preferences, the same reasoning applies.  $\square$

## References

- Blackorby, C., Donaldson, D., 1992. Pigs and guinea pigs: A note on the ethics of animal exploitation. *The Economic Journal* 102, 1345–1369.
- Decerf, B., Van der Linden, M., 2015. Fair social orderings with other-regarding preferences. *Social Choice and Welfare* 46, 655–694.
- Fleurbaey, M., Maniquet, F., 2011. *A Theory of Fairness and Social Welfare*. Cambridge University Press.
- Fleurbaey, M., Trannoy, A., 2003. The impossibility of a Paretian egalitarian. *Social Choice and Welfare* 21, 243–263.
- Hammond, P.J., 1976. Equity, Arrow's conditions, and Rawls' difference principle. *Econometrica* 44, 793–804.

- Hansson, B., 1973. The independence condition in the theory of social choice. *Theory and Decision* 4, 25–49.
- Kymlicka, W., Donaldson, S., 2011. *Zoopolis: A Political Theory of Animal Rights*. Oxford University Press Oxford, Oxford.
- Maniquet, F., Sprumont, Y., 2004. Fair production and allocation of a non-rival good. *Econometrica* 72, 627–640.
- Maniquet, F., Sprumont, Y., 2005. Welfare egalitarianism in nonrival environments. *Journal of Economic Theory* 120, 155–174.
- Moulin, H., 1987. Egalitarian-equivalent cost sharing of a public good. *Econometrica* 55, 963–976.
- Singer, P., 1975. *Animal Liberation*. HarperCollins, New York.
- Sprumont, Y., 2012. Resource egalitarianism with a dash of efficiency. *Journal of Economic Theory* 147, 1602–1613.
- Treibich, R., 2019. Welfare egalitarianism with other-regarding preferences. *Social Choice and Welfare* 52, 1–28.
- Van der Linden, M., 2018. Egalitarianism with a dash of fair efficiency. *Economic Theory Bulletin* Forthcoming.

## Acknowledgments

This paper was previously circulated under the title “If you’re an egalitarian, how come you eat meat?”. We are grateful to

ADD

Rafael Treibich and an anonymous reviewer for helpful suggestions and comments.

We also thank participants to presentations at

ADD

the Hoover Chair, Universite Catholique de Louvain Universit e Catholique de Louvain and the conference on “Animals and Social Welfare” at Duke University for useful questions and discussions.