

# An Axiomatic Approach to Revising Preferences

Adrian Haret<sup>1</sup>, Johannes P. Wallner<sup>2</sup>

<sup>1</sup> Institute for Logic, Language and Computation (ILLC), University of Amsterdam, The Netherlands

<sup>2</sup> Institute of Software Technology, Graz University of Technology, Austria  
a.haret@uva.nl, wallner@ist.tugraz.at

## Abstract

We study a model of preference revision in which a prior preference over a set of alternatives is adjusted in order to accommodate input from an authoritative source, while maintaining certain structural constraints (e.g., transitivity, completeness), and without giving up more information than strictly necessary. We analyze this model under two aspects: the first allows us to capture natural distance-based operators, at the cost of a mismatch between the input and output formats of the revision operator. Requiring the input and output to be aligned yields a second type of operator, which we characterize using preferences on the comparisons in the prior preference. Preference revision is set in a logic-based framework and using the formal machinery of belief change, along the lines of the well-known AGM approach: we propose rationality postulates for each of the two versions of our model and derive representation results, thus situating preference revision within the larger family of belief change operators.

## 1 Introduction

Preferences play a central role in theories of decision making as part of the mechanism underlying rational choice: they show up in economic models of rational agency (Sen 2017), as well as in formal models of artificial agents expected to interact with the world and each other (Domshlak et al. 2011; Rossi, Venable, and Walsh 2011; Pigozzi, Tsoukiàs, and Viappiani 2016). Since such interactions take place in dynamic environments, it can be expected that preferences change in response to new developments.

In this paper we are interested in preference change occurring when new preference information becomes available and has to be taken at face value, thereby prompting a change in the prior preference. The change, we require, should preserve as much useful information from the prior preference as can be afforded. Preference change thus described is a pervasive phenomenon, arising in many contexts spanning the realms of both human and artificial agency. One prominent example is the distinguished tradition in Economics and Philosophy looking at examples of conflict between an agent’s subjective preference (what we call here the prior preference  $\pi$ ) and a second-order preference, often standing for a commitment or moral rule (what we call

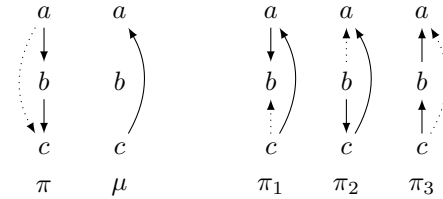


Figure 1: Revising  $\pi$  by  $\mu$  can be thought of a choice between which comparisons to keep and which to give up.

here the new preference information  $\mu$ ): subjective versus ‘ethical’ preferences (Harsanyi 1955), lack of will, or *akrasia* (Jeffrey 1974), moral commitments (Sen 1977), second-order volitions (Frankfurt 1988) and second-order preferences (Nozick 1994) all fall under this heading.

The same challenge can occur in technological applications, from updating CP-nets (Cadilhac et al. 2015) to changing the order in which search results are displayed on a page in response to user provided specifications, as well as, more generally, in issues related to the *alignment problem* (Russell 2019): an artificial agent dealing with humans will have to learn their preferences, but as it cannot do so instantaneously, it must presumably do so in intermediate steps, revising along the way. The following example illustrates the problem in its most basic form.

**Example 1.** *An online streaming service constructs a profile tailored to a particular user, according to which the arthouse movie (a) is preferred to the biopic (b), which is preferred to the comedy (c), and thus displays them in this order, encoded here with the preference statement  $\pi = (a \succ b) \wedge (b \succ c)$ . When the user volunteers information to the effect that they find the comedy better than the arthouse movie, i.e., new information  $\mu = (c \succ a)$ , the streaming service must revise its model of the user’s preference: it has to place c before a and, in order to display the alternatives in a neat linear fashion, it must decide on a position to slot b into. Preferences  $\pi$  and  $\mu$ , together with possible values for the revised result, e.g.,  $\pi_1 = (c \succ a) \wedge (a \succ b)$ ,  $\pi_2 = (c \succ a) \wedge (b \succ c)$  and  $\pi_3 = (c \succ a) \wedge (c \succ b) \wedge (b \succ a)$ , are depicted in Figure 1. Intuitively,  $\pi_3$  veers far (too far) away from the input preference  $\pi$ , in that it does not keep any of the still permissible comparisons contained in  $\pi$ , and should arguably be excluded,*

while  $\pi_1$  and  $\pi_2$  are viable contenders. If we go further and insist on a decision between  $\pi_1$  and  $\pi_2$ , we can take stock of the information relayed by either choice. Accepting  $\pi_1$  involves giving up the comparison of  $b$  over  $c$ , and we may surmise this is because the comparison of  $a$  over  $b$  is given up more reluctantly: preference of the arthouse movie over the biopic is more intense! Acceptance of  $\pi_2$  implies the opposite:  $b$  over  $c$  is now the stronger preference. Thus, restricting the output of revision to a single linear order suggests that the choice can be rationalized using an implicit preference order over the comparisons.

Thus, whether it is the internal conflict of a moral agent or a content provider aiming for a better user experience, many cases of preference change involve a conflict between two types of preferences, one of which has priority. But, despite the fact that the problem is often signaled, a principled approach to how to handle it is often overlooked.

Our purpose here is to formalize the type of reasoning illustrated in Example 1 by rationalizing preference change as a type of choice function that utilizes the information provided by the prior preference in adapting itself to new information. In particular, we want to combine techniques from standard belief change with Sen’s insight that conflicts among preferences should be resolved using preferences over the preferences themselves (Sen 1977).

**Contributions.** We put forward two models for preference revision. The first, called here *irresolute revision*, is based on minimizing distance to the input preference, expressed as a formula, while allowing the output to be represented as a set of formulas, and is the approach that follows most closely the one from propositional revision. However, in an interesting departure from the propositional model, we show that insistence on representing the revised preference as a single formula, what we call *resolute revision*, requires us to give up appealing distance-based operators. The second model picks up where this result leaves off, and we present a mechanism for resolute revision based on an underlying preference relation over the adjacent comparisons of the prior preference, construed here as atomic elements that can be subject to change. In both cases we present desirable normative principles, in the AGM mould (Alchourrón, Gärdenfors, and Makinson 1985), and derive representation results.

**Related work.** Previous work dealing with preference revision has studied preference change prompted by a change in beliefs (Bradley 2007; Lang and van der Torre 2008; Liu 2011). Here we abstract away from the source of new information, focusing exclusively on a mechanism for resolving conflicts between preferences. Other work (Cadilhac et al. 2015) describes preference change when preferences are represented using CP-nets (Boutilier et al. 2004), or dynamic epistemic logic (Bentham and Liu 2014), in the context of declarative debugging (Dell’Acqua and Pereira 2005), or databases (Chomicki 2003), and therefore comes with additional structural constraints. The basic phenomenon of preference change has also been raised in explicit connection to belief change (Hansson 1995; Grüne-Yanoff and Hansson 2009; Grüne-Yanoff 2013), but a representation in terms of preferences on the comparisons present in the preference

orders, along the lines suggested here, has, to the best of our knowledge, not yet been given. Much existing work proceeds by putting forward some concrete preference revision mechanism, occasionally with a remark on the similarity to belief revision (Freund 2004; Chomicki and Song 2005; Liu 2011; Ma, Benferhat, and Liu 2012). Our work adds an analysis in terms of postulates and representation results.

The framework studied here is inspired by propositional belief revision (Alchourrón, Gärdenfors, and Makinson 1985; Katsuno and Mendelzon 1992; Nebel 1992), and the two problems are similar in spirit. The closest analogy for Section 3 is belief revision in fragments of propositional logic (Delgrande, Peppas, and Woltran 2018), while the equivalent for Section 4 would be an attempt to characterize propositional revision operators using orders on the atoms, with similar results derived in (Haret and Wallner 2021).

**Outline.** In Section 2 we introduce notation and the basic elements of our model. Section 3 looks at irresolute revision operators and Section 4 looks at resolute revision. Section 5 offers concluding remarks.

## 2 Preliminaries

We assume a finite set  $A$  of alternatives, with  $|A| = n$ . For alternatives  $x, y \in A$ , we call the ordered pair  $(x, y)$  a *comparison*, and think of it as encoding the fact that the agent prefers  $x$  to  $y$ . For ease of reading we write  $xy$  instead of  $(x, y)$ . A (strict) *linear order*  $\ell$  on  $A$  is a binary relation (i.e., a set of comparisons) on  $A$  such that (i)  $xx \notin \ell$ , for any  $x \in A$  (irreflexivity), (ii) if  $xy \in \ell$  and  $yz \in \ell$  then  $xz \in \ell$  (transitivity), (iii) for any distinct  $x, y \in A$ , either  $xy \in \ell$  or  $yx \in \ell$  (connectedness). We write  $\ell$  as the string  $x_1 \dots x_n$ , where  $x_i x_j \in \ell$ , for  $i < j$ , and denote by  $\mathcal{L}_A$  the set of linear orders over  $A$ . A *strict partial order*  $\delta$  on  $A$  is an irreflexive and transitive (not necessarily complete) binary relation on  $A$ . Note that irreflexivity and transitivity imply that if  $xy \in \delta$  then  $yx \notin \delta$  (antisymmetry). Note, also, that if  $\ell_1$  and  $\ell_2$  are linear orders on  $A$ , then  $\ell_1 \cap \ell_2$  is *not* guaranteed to be a linear order, though it is a strict partial order on  $A$ .

An *atomic preference statement* has the form  $x \succ y$ , for distinct alternatives  $x, y \in A$ . A *preference statement*  $\pi$  is a conjunction of atomic preference statements, with  $\mathcal{P}_A$  as the set of all preference statements over  $A$ . A linear order  $\ell$  satisfies the atomic preference statement  $x \succ y$  if  $xy \in \ell$ ;  $\ell$  satisfies a preference statement  $\pi$  if it satisfies every atomic preference statement in  $\pi$ , in which case we say that  $\ell$  is a *model* of  $\pi$ . The *set*  $[\pi]$  of models of  $\pi$  is defined as  $[\pi] = \{\ell \in \mathcal{L}_A \mid \ell \text{ satisfies } \pi\}$ . If  $\Pi = \{\pi_1, \dots, \pi_n\}$  is a set of preference statements, the *set*  $[\Pi]$  of models of  $\Pi$  is defined as  $[\Pi] = \bigcup_{1 \leq i \leq n} [\pi_i]$ , i.e., as the union of the models of the formulas in  $\Pi$ . A preference statement  $\pi$  (set of preference statements  $\Pi$ ) is *consistent* if  $[\pi] \neq \emptyset$  ( $[\Pi] \neq \emptyset$ ).

**Example 2.** For  $A = \{a, b, c\}$ ,  $\pi = (a \succ b) \wedge (a \succ c)$  is a preference statement indicating that  $a$  is preferred, once, to  $b$  and, second, to  $c$ . The set of models of  $\pi$  is  $[\pi] = \{abc, acb\}$ . Note that  $abc \cap acb = \{ab, ac\}$ , i.e., the intersection of  $abc$  and  $acb$  is the partial order containing the comparisons that  $abc$  and  $acb$  have in common.

### 3 Irresolute Preference Revision

An *irresolute preference revision operator*  $\circ$  is a function  $\circ: \mathcal{P}_A \times \mathcal{P}_A \rightarrow 2^{\mathcal{P}_A}$ , taking as input two preference statements, typically denoted  $\pi$  and  $\mu$ , and standing for the agent's prior and newly acquired preference information, respectively, and returning a *set* of preference statements, denoted  $\pi \circ \mu$ . The representation of the result as a set of formulas is a slight departure from established revision practice, but has precedent in belief change applied to formalisms other than propositional logic, e.g., in work on the aggregation of abstract Argumentation Frameworks (Delobelle et al. 2016). Intuitively,  $\pi \circ \mu$  can be interpreted as a range of options, all of which, together, represent the agent's adjusted preferences in light of new information  $\mu$ . We will have occasion to reflect on the format of the result later on.

In typical revision fashion, we want to identify a set of desirable normative principles that (irresolute) preference revision operators ideally satisfy. To this purpose we use an adapted version of the well-known AGM postulates (Alchourrón, Gärdenfors, and Makinson 1985), or rather, their KM formulation (Katsuno and Mendelzon 1992):

- (R<sub>1</sub><sup>i</sup>)  $[\pi \circ \mu] \subseteq [\mu]$ .
- (R<sub>2</sub><sup>i</sup>) If  $\pi \wedge \mu$  is consistent, then  $\pi \circ \mu = \{\pi \wedge \mu\}$ .
- (R<sub>3</sub><sup>i</sup>) If  $\mu$  is consistent, then  $\pi \circ \mu$  is consistent.
- (R<sub>4</sub><sup>i</sup>) If  $[\pi_1] = [\pi_2]$  and  $[\mu_1] = [\mu_2]$ , then  $[\pi_1 \circ \mu_1] = [\pi_2 \circ \mu_2]$ .
- (R<sub>5</sub><sup>i</sup>)  $[\pi \circ \mu_1] \cap [\mu_2] \subseteq [\pi \circ (\mu_1 \wedge \mu_2)]$ .
- (R<sub>6</sub><sup>i</sup>) If  $[\pi \circ \mu_1] \cap [\mu_2] \neq \emptyset$ , then  $[\pi \circ (\mu_1 \wedge \mu_2)] \subseteq [\pi \circ \mu_1] \cap [\mu_2]$ .

Even though the underlying semantics of the formulas concerns linear orders, the intuition and mechanics behind the postulates is entirely similar to that of propositional revision, and we direct the reader to standard references for more details (Alchourrón, Gärdenfors, and Makinson 1985; Katsuno and Mendelzon 1992; Fermé and Hansson 2018). Owing to the blend of different formats we write the axioms in their semantic version (i.e., using the sets of models), rather than in the more familiar syntactic formulation (i.e., using entailment relations and conjunction operators), but their motivation is otherwise unchanged from the propositional case.

It turns out, as expected, that axioms R<sub>1</sub><sup>i</sup>–R<sub>6</sub><sup>i</sup> characterize a broad class of revision operators, one that can be described using the familiar device of total pre-orders (i.e., complete, transitive binary relations) on the set  $\mathcal{L}_A$  of linear orders and the notion of a *preference assignment*  $f$ , i.e., a function  $f: \mathcal{P}_A \rightarrow \mathcal{T}_A$ , where  $\mathcal{T}_A$  is the set of total preorders on  $\mathcal{L}_A$ . A preference assignment maps a preference statement  $\pi$  to a total preorder on linear orders, denoted here as  $\leq_\pi$ : the preferences on preferences mentioned in Section 1. Using established revision lingo (Katsuno and Mendelzon 1992), a preference assignment is *faithful* if it satisfies the following properties:

- (f<sub>1</sub><sup>i</sup>) If  $\ell_1, \ell_2 \in [\pi]$ , then  $\ell_1 \approx_\pi \ell_2$ .
- (f<sub>2</sub><sup>i</sup>) If  $\ell_1 \in [\pi]$  and  $\ell_2 \notin [\pi]$ , then  $\ell_1 <_\pi \ell_2$ .
- (f<sub>3</sub><sup>i</sup>) If  $[\pi_1] = [\pi_2]$ , then  $\ell_1 \leq_{\pi_1} \ell_2$  iff  $\ell_1 \leq_{\pi_2} \ell_2$ .

A faithful preorder  $\leq_\pi$  makes the models of  $\pi$  as the uniquely most preferred elements of  $\leq_\pi$ , regardless of the syntax of  $\pi$ . The characterization, then, proceeds as follows.

**Theorem 1.** *An irresolute revision operator  $\circ$  satisfies postulates R<sub>1</sub><sup>i</sup>–R<sub>6</sub><sup>i</sup> iff there exists a faithful preference assignment mapping every preference statement  $\pi$  to a total preorder  $\leq_\pi$  such that  $[\pi \circ \mu] = \min_{\leq_\pi} [\mu]$ .*

The proof of Theorem 2 is analogous to its equivalent version in propositional revision (Katsuno and Mendelzon 1992), with one notable exception, on which more shortly.

An important device for generating concrete revision operators is a *distance*  $d$  between linear orders, i.e., a function  $d: \mathcal{L}_A \times \mathcal{L}_A \rightarrow \mathbb{R}_{\geq 0}$ , such that  $d(\ell_1, \ell_2) = 0$  if and only if  $\ell_1 = \ell_2$ . This is a minimal requirement on  $d$ , on top of which we will add more desirable properties later on. A typical distance we will use here is the *Kendall tau distance*  $d_\tau$  (Kendall and Gibbons 1990) defined as  $d_\tau(\ell_1, \ell_2) = |\{xy \in \ell_1 \mid yx \in \ell_2\}|$ , i.e., as the number of disagreements (inverted pairs of alternatives) between  $\ell_1$  and  $\ell_2$ . Less discriminating, the *drastic distance*  $d_D$  is defined as  $d_D(\ell_1, \ell_2) = 0$ , if  $\ell_1 = \ell_2$ , and  $k > 0$ , otherwise. Given a distance  $d$ , the *d-induced irresolute preference revision operator*  $\circ^d$  is defined, for any preference statements  $\pi$  and  $\mu$ , as a set of preference statements  $\pi \circ^d \mu$  such that:

$$[\pi \circ^d \mu] = \operatorname{argmin}_{\ell \in [\mu]} \min_{\ell' \in [\pi]} d(\ell', \ell),$$

i.e., a set of preference statements whose models add up to exactly those models of  $[\mu]$  that minimize the Kendall tau distance to any model of  $\pi$ . We write  $\circ^\tau$  and  $\circ^D$  for the  $d_\tau$ - and  $d_D$ -induced revision operators, respectively. Note that  $\circ^D$  can be written as:

$$\pi \circ^D \mu = \begin{cases} \{\pi \wedge \mu\}, & \text{if } \pi \wedge \mu \text{ is consistent,} \\ \{\mu\}, & \text{otherwise.} \end{cases}$$

Importantly, note that  $d$ -induced revision operators are well-defined, as any individual linear order  $\ell = x_1 \dots x_n$  is the sole model of the formula  $\pi_\ell = (x_1 \succ x_2) \wedge \dots \wedge (x_{n-1} \succ x_n)$ , and any set  $\{\ell_1, \dots, \ell_m\}$  of linear orders is the set of models of the set  $\{\pi_{\ell_1}, \dots, \pi_{\ell_m}\}$  of preference statements.

**Example 3.** *For  $A = \{a, b, c\}$  and  $\pi = (a \succ b) \wedge (b \succ c)$ ,  $\mu = (c \succ a)$ , we have that  $[\pi] = \{abc\}$  and  $[\mu] = \{cab, cba, bca\}$ . The Kendall tau distances between the model of  $\pi$  and the models of  $\mu$  are  $d_\tau(abc, cab) = 2$ ,  $d_\tau(abc, cba) = 3$ ,  $d_\tau(abc, bca) = 2$ , and thus  $[\pi \circ^\tau \mu] = \{cab, bca\}$ . We can represent  $[\pi \circ^\tau \mu]$  using preference statements  $\pi_1 = (c \succ a) \wedge (a \succ b)$  and  $\pi_2 = (b \succ c) \wedge (c \succ a)$ , noting that  $[\pi_1] \cup [\pi_2] = \{cab\} \cup \{bca\} = \{cab, bca\}$ , with  $\pi \circ^\tau \mu = \{\pi_1, \pi_2\}$ .*

Given a distance function  $d$  we can define the *d-induced preorder*  $\leq_\pi^d$  by taking  $\ell_1 \leq_\pi^d \ell_2$  if  $\min_{\ell \in [\pi]} d(\ell, \ell_1) \leq \min_{\ell \in [\pi]} d(\ell, \ell_2)$ , and it is straightforward to see that the  $d_\tau$ - and  $d_D$ -induced preorders  $\leq_\pi^\tau$  and  $\leq_\pi^D$ , respectively, are faithful; this, via Theorem 1, implies that  $\circ^\tau$  and  $\circ^D$  satisfy axioms R<sub>1</sub><sup>i</sup>–R<sub>6</sub><sup>i</sup>. Throughout all this, though, a key detail is the fact that for any set  $L$  of linear orders we can find, as described earlier, a set  $\Pi$  of preference statements such that  $[\Pi] = L$ . At the beginning of this section we offered an intuition of what  $\Pi$  stands for, but we may now add to a concern about this design choice, as it introduced a mismatch between the input and output. This makes it harder, for instance, to apply a revision operator iteratively. It would be

desirable, therefore, to represent  $[\pi \circ \mu]$  as a *single* preference statement, rather than as a set. However, as Proposition 1 below shows, this is possible only in special circumstances. To state Proposition 1, we introduce one additional piece of notation. If  $\delta$  is a strict partial order, the linear order  $\ell$  is a *completion* of  $\delta$  if  $\delta \subseteq \ell$ , i.e., if  $\ell$  preserves all the comparisons in  $\delta$ . We denote by  $\text{comp}(\delta)$  the set of completions of  $\delta$ .

**Proposition 1.** *If  $L \subseteq \mathcal{L}_A$ , then there exists  $\pi \in \mathcal{P}_A$  such that  $[\pi] = L$  iff  $L = \text{comp}(\bigcap_{\ell \in L} \ell)$ .*

*Proof.* Note that  $\bigcap_{\ell \in L} \ell$  is a strict partial order, which we denote by  $\delta_L$ . Thus, if  $L = \text{comp}(\bigcap_{\ell \in L} \ell)$ , then  $\pi = \bigwedge_{xy \in \delta_L} (x \succ y)$  is the preference statement we are looking for, as  $[\pi] = L$ . Conversely, suppose there exists some  $\pi \in \mathcal{P}_A$  such that  $[\pi] = L$ . Take, first, a linear order  $\ell \in \text{comp}(\bigcap_{\ell \in L} \ell)$ : we claim that  $\ell \in [\pi]$ . To see why this is the case, assume that  $\ell \notin [\pi]$ : this implies that there exists a comparison  $xy \in \ell$ , such that its opposite  $yx$  is implied by  $\pi$ . The latter fact implies that  $yx \in \ell'$ , for every  $\ell' \in [\pi]$ , which further implies that  $yx \in \bigcap_{\ell \in L} \ell$ . Thus  $yx \in \ell''$ , for every  $\ell'' \in \text{comp}(\bigcap_{\ell \in L} \ell)$ : *a fortiori*,  $yx \in \ell$ , which is a contradiction. We have thus obtained that  $\text{comp}(\bigcap_{\ell \in L} \ell) \subseteq L$ .

Next, take  $\ell \in L$ . Since  $\ell$  satisfies every comparison in  $\pi$ , we get that  $\bigcap_{\ell' \in L} \ell' \subseteq \ell$ , and thus  $\ell \in \text{comp}(\bigcap_{\ell' \in L} \ell')$ , showing that  $L \subseteq \text{comp}(\bigcap_{\ell \in L} \ell)$ .  $\square$

Thus, Proposition 1 shows that a set  $L$  of linear orders can be encoded by a preference statement  $\pi$  only if  $L$  satisfies what we may think of as a closure operation:  $L$  must be closed under completions of the strict partial order that contains the comparisons common amongst all the orders in  $L$ .<sup>1</sup> Coming back to the issue of whether outputs of distance-based irresolute preference revision operators can be squeezed into  $\mathcal{P}_A$ , we see that this result spells trouble.

**Example 4.** *For  $A$ ,  $\pi$  and  $\mu$  as in Example 3, we get that  $[\pi \circ^\tau \mu] = \{cab, bca\}$ . Note that  $cab \cap bca = \{ca\}$ , and  $\text{comp}(\{ca\}) = \{cab, cba, bca\} \neq [\pi \circ^\tau \mu]$ . Thus, using Proposition 1, there is no preference formula  $\pi' \in \mathcal{P}_A$  such that  $[\pi \circ^\tau \mu] = [\pi']$ .*

In other words, we cannot have the output of a preference revision operator be a single preference statement *and*, at the same time, hold on to  $\circ^\tau$ . And it turns out that, under mild assumptions on the distance function  $d$ , this situation is unavoidable for any distance-based revision operator  $\circ^d$ . To state these properties, we introduce a number of new notions. A *renaming*  $r$  is a bijective function  $r: A \rightarrow A$ , with the renaming  $r(\ell)$  of  $\ell = x_1 \dots x_n$  defined as  $r(x_1) \dots r(x_n)$ . For linear orders  $\ell_1 = x_1 \dots x_m$  and  $\ell_2 = x_{m+1} \dots x_n$  on distinct alternatives, the *concatenation*  $\ell_1 \ell_2$  of  $\ell_1$  and  $\ell_2$  is the linear order defined as  $\ell_1 \ell_2 = x_1 \dots x_m x_{m+1} \dots x_n$ . The properties we expect from the distance function  $d$  are:

- (D<sub>1</sub>)  $d(\ell_1, \ell_2) = d(\ell_2, \ell_1)$ .
- (D<sub>2</sub>)  $d(\ell_1, \ell_2) = d(r(\ell_1), r(\ell_2))$ .
- (D<sub>3</sub>) If  $d_\tau(\ell_1, \ell_2) < d_\tau(\ell_1, \ell_3)$ , then  $d(\ell_1, \ell_2) < d(\ell_1, \ell_3)$ .
- (D<sub>4</sub>)  $d(\ell \ell_1 \ell', \ell \ell_2 \ell') = d(\ell_1, \ell_2)$ .

Property D<sub>1</sub> requires  $d$  to be symmetric; D<sub>2</sub> requires  $d$  to be invariant under renamings; D<sub>3</sub> is a monotonicity condition requiring the distance function  $d$  to be consistent with the tau distance  $d_\tau$ , i.e., the more adjacent pairs of alternatives we flip, starting with  $\ell_1$ , the further away from  $\ell_1$  we end up; D<sub>4</sub> requires the distance between  $\ell_1$  and  $\ell_2$  to depend only on the distinct sections of  $\ell_1$  and  $\ell_2$ . Finally, we say that an (irresolute) preference revision operator  $\circ$  is *single-statement compliant* if for any set  $A$  of alternatives and preference statements  $\pi, \mu \in \mathcal{P}_A$ , there exists a preference statement  $\pi' \in \mathcal{P}_A$  such that  $[\pi \circ \mu] = [\pi']$ . Under these conditions, we can prove the following result.

**Theorem 2.** *If a distance function  $d$  satisfies properties D<sub>1</sub>–D<sub>4</sub>, the operator  $\circ^d$  is not single-statement compliant.*

*Proof.* Assume  $\circ^d$  is single-statement compliant and take  $A = \{a, b, c\}$ . Using the renaming  $r(a) = c$ ,  $r(b) = a$  and  $r(c) = b$ , we have that:

$$\begin{aligned} d(abc, bca) &= d(cab, abc) && \text{by D}_2, \text{ applying } r \\ &= d(abc, cab). && \text{by D}_1 \end{aligned}$$

By property D<sub>3</sub>, it holds that  $d(abc, bca) < d(abc, cba)$  and  $d(abc, cab) < d(abc, cba)$ . Consider preference statements  $\pi = (a \succ b) \wedge (b \succ c)$  and  $\mu = (c \succ a)$  with  $[\pi] = \{abc\}$  and  $[\mu] = \{cab, cba, bca\}$ . Using the just derived distance relationships we infer that  $[\pi \circ^d \mu] = \{bca, cab\}$ . By the assumption that  $\circ^d$  is single-statement compliant there must exist some preference statement  $\pi' \in \mathcal{P}_A$  such that  $[\pi'] = \{bca, cab\}$ . By Proposition 1, however, this leads to a contradiction. Using property D<sub>4</sub>, we can extend this result to any alphabet with at least three elements.  $\square$

Note that  $d_\tau$  satisfies properties D<sub>1</sub>–D<sub>4</sub>, while  $d_D$  satisfies all but D<sub>3</sub> and thus does not fall within the scope of Theorem 2. However,  $d_D$  does satisfy a weaker version of D<sub>3</sub> where the inequality is non-strict, and a slightly more involved argument shows that  $d_D$  is the only distance function that satisfies these properties and is single-statement compliant for three alternatives. In fact, we conjecture that  $\circ^D$  is the only irresolute preference revision operator that is single-statement compliant, for *any* number of alternatives.

## 4 Resolute Preference Revision

In the wake of Section 3, we want to understand how to align the formats of the prior and revised preference information. In this section we show that this is possible by assuming that the user has some preference structure on the atomic elements of its prior preference. We focus on the case in which both prior and revised preferences encode single linear orders, which makes sense in light of Example 1, where the alternatives need to be sequentially displayed, e.g., on a webpage. As users cannot be expected to be exhaustive in the information they provide, we continue to allow  $\mu$  to be a regular (not necessarily complete) preference statement.

<sup>1</sup>This issue does not show up in propositional revision, where any set of truth-value assignments can be represented by a formula, but it does occur in belief change in fragments of propositional logic (Delgrande, Peppas, and Woltran 2018), or with respect to Argumentation Frameworks (Diller et al. 2015; Haret, Wallner, and Woltran 2018).

A preference statement is *complete* if it has exactly one model, and we denote by  $\mathcal{C}_A$  the set of complete preference statements on  $A$ . A *resolute preference revision operator*  $\circ$  is a function  $\circ: \mathcal{C}_A \times \mathcal{P}_A \rightarrow \mathcal{C}_A$ , taking as input a complete and a standard preference statement, typically denoted by  $\lambda$  and  $\mu$ , respectively, and returning a complete preference statement, denoted by  $\lambda \circ \mu$ . Since  $\lambda \circ \mu$  has, by definition, exactly one linear order  $\ell$  as model, we often use  $\ell$  and  $\{\ell\}$  interchangeably.

The road to representation starts by laying down some additional items of notation. If  $C$  is a set of comparisons on  $A$ , the *transitive closure*  $C^+$  of  $C$  is the smallest set such that (i)  $C \subseteq C^+$  and (ii) if  $xy, yz \in C^+$ , then  $xz \in C^+$ . Intuitively,  $C^+$  is a transitive relation on  $A$  that may, or may not, contain cycles between alternatives. If  $\mu$  is a preference statement on  $A$ , the *set*  $C_\mu$  of comparisons of  $\mu$  is defined as  $C_\mu = \{xy \in A^2 \mid x \succ y \text{ is an atomic preference in } \mu\}$ , i.e.,  $C_\mu$  contains the comparisons explicitly sanctioned by  $\mu$ , while  $C_\mu^+$  adds the comparisons inferred by transitivity. For a complete preference statement  $\lambda$ ,  $C_\lambda^+$  is, by definition, a linear order: the unique model of  $\lambda$ .

If  $[\lambda] = \{x_1 \dots x_n\}$ , then  $x_i x_{i+1}$ , for  $1 \leq i \leq n-1$  is an *adjacent comparison* of  $\lambda$ , with  $A_\lambda$  being the set of adjacent comparisons of  $\lambda$ . Adjacent comparisons are the basic atoms we will take as the basis for the preference relations guiding resolute revision operators. If  $\mu$  is a preference statement, a  $\lambda$ -model of  $\mu$  is a linear order  $\ell^*$  such that  $\ell^* = (C_\mu \cup C)^+$  for some set of comparisons  $C \subseteq C_\lambda^+$ . Intuitively, a  $\lambda$ -model of  $\mu$  is a linear order obtained by adding to  $\mu$  some of the comparisons expressed, either explicitly or implicitly, by  $\lambda$ . We write  $[\mu]_\lambda$  for the set of  $\lambda$ -models of  $\mu$ . Note that  $[\mu]_\lambda \neq \emptyset$  if  $[\mu] \neq \emptyset$ : if the order between two alternatives  $x$  and  $y$  is not decided by  $\mu$ , then we can complete the order by reaching into  $C_\lambda^+$ , since  $C_\lambda^+$  is a linear order and either  $xy \in C_\lambda^+$  or  $yx \in C_\lambda^+$ . If  $\lambda \in \mathcal{C}_A$  and  $\mu \in \mathcal{P}_A$ , the *cycle-free part*  $CF_\mu(\lambda)$  of  $\lambda$  with respect to  $\mu$  is defined as  $CF_\mu(\lambda) = \{xy \in C_\lambda^+ \mid xy \text{ is not involved in a cycle in } C_{\lambda \wedge \mu}^+\}$ , i.e., the comparisons in  $\lambda$  (both explicit and implicit) that are not part of a cycle in the relation obtained when adding  $\mu$  to  $\lambda$ .

**Example 5.** Take  $\lambda = (a \succ b) \wedge (b \succ c) \wedge (c \succ d)$ ,  $\mu = (c \succ a)$ , with  $[\lambda] = \{abcd\}$ ,  $[\mu] = \{cabd, cbad, bcad, cadb, \dots\}$ ,  $C_\lambda^+ = \{ab, ac, ad, bc, bd, cd\}$ ,  $C_\mu = C_\mu^+ = \{ca\}$ . The adjacent comparisons of  $\lambda$  are  $A_\lambda = \{ab, bc, cd\}$ . The conjunction of  $\lambda \wedge \mu$  contains a cycle between  $a$ ,  $b$  and  $c$ , hence  $CF_\mu(\lambda) = \{ad, bd, cd\}$ . Note, next, that  $cabd = (C_\mu \cup CF_\mu(\lambda) \cup \{ab\})^+$ , i.e., it is the linear order obtained by adding the comparison  $ab \in abcd$  as well as the cycle-free part of  $\lambda$  to  $ca$ , and then taking the transitive closure of the resulting set of comparisons. Similarly,  $bcad$  is obtained by adding  $bc \in abc$  to  $C_\mu$  and  $CF_\mu(\lambda)$ . Adding to  $C_\mu$  any set of comparisons from  $abc$  that includes  $ac$  yields a cycle between  $a$ ,  $b$  and  $c$  and is therefore not fit to construct a linear order. Adding  $\emptyset$  to  $C_\mu$  is also no good, as the result is not complete. Thus,  $[\mu]_\lambda = \{cabd, bcad\}$ .

The notion of a  $\lambda$ -model is important in allowing us to formulate desirable axioms for resolute preference revision:

(R<sub>1</sub><sup>r</sup>) If  $\mu$  is consistent, then  $[\lambda \circ \mu] \subseteq [\mu]_\lambda$ .

(R<sub>2</sub><sup>r</sup>) If  $\mu$  is inconsistent, then  $\lambda \circ \mu = \lambda$ .

(R<sub>3</sub><sup>r</sup>) If  $[\lambda_1] = [\lambda_2]$  and  $[\mu_1] = [\mu_2]$ , then  $[\lambda_1 \circ \mu_1] = [\lambda_2 \circ \mu_2]$ .

(R<sub>4</sub><sup>r</sup>) If  $(\lambda \circ \mu_1) \wedge \mu_2$  is consistent, then  $[(\lambda \circ \mu_1) \wedge \mu_2] = [\lambda \circ (\mu_1 \wedge \mu_2)]$ .

Though built on familiar intuitions, axioms R<sub>1</sub><sup>r</sup>–R<sub>4</sub><sup>r</sup> are particular enough to warrant discussion. Note, first, that  $\lambda \circ \mu$  being a preference statement allows us to connect it to other preference statements *via* the conjunction operator  $\wedge$ . Then, given that  $\lambda \circ \mu$  is a complete statement by definition, R<sub>1</sub><sup>r</sup> requires it to define a  $\lambda$ -model of  $\mu$ , i.e., to be obtained using only comparisons from  $\lambda$  in addition to those of  $\mu$ : a restriction meant to prevent  $\lambda \circ \mu$  from introducing comparisons not justified by their presence in  $\lambda$ , as illustrated below.

**Example 6.** For  $\lambda$  and  $\mu$  as in Example 5 we obtained  $[\mu]_\lambda = \{cabd, bcad\}$ . The linear order  $cbad \in [\mu]_\lambda$  is not desirable as a candidate for  $\lambda \circ \mu$ , as it is obtained by adding  $cb$  and  $ba$  to  $ca$  (in addition to the cycle-free comparisons); however, neither  $cb$  nor  $ca$  are in  $abcd$ , i.e., their addition is unjustified based on the prior preference information.

Axiom R<sub>2</sub><sup>r</sup> ensures that  $\lambda \circ \mu$  is defined even when  $\mu$  contains a cycle. Together, axioms R<sub>1</sub><sup>r</sup> and R<sub>2</sub><sup>r</sup> have the additional desirable effect of ensuring that  $\lambda \circ \mu$  holds on to any comparisons of  $\lambda$  not involved in a cycle with  $\mu$ , or not explicitly ruled out by  $\mu$ .

**Proposition 2.** If  $\lambda \in \mathcal{C}_A$ ,  $\mu \in \mathcal{P}_A$ , and  $xy \in C_\lambda^+$  such that  $yx \notin C_\mu^+$  and there is no cycle involving  $xy$  in  $C_{\lambda \wedge \mu}^+$ , then, if  $\circ$  is a resolute preference revision operator satisfying axioms R<sub>1</sub><sup>r</sup> and R<sub>2</sub><sup>r</sup>, it holds that  $xy \in C_{\lambda \circ \mu}^+$ .

*Proof.* Assuming that  $xy \notin C_{\lambda \circ \mu}^+$ , we conclude, by completeness of  $C_{\lambda \circ \mu}^+$ , that  $yx \in C_{\lambda \circ \mu}^+$ . By axiom R<sub>1</sub><sup>r</sup>, we know that  $C_{\lambda \circ \mu}^+ = (C_\mu \cup C)^+$ , for some set of comparisons in  $C_\lambda^+$ . We cannot have that  $yx \in C_\mu$ , or even that  $yx \in C_\mu^+$ , as this would contradict our assumptions. It must be the case, then, that  $yx$  is obtained as the transitive closure of comparisons  $yz_1, x_1 z_2, \dots, y z_t$ , with some comparisons coming from  $\lambda$  and others from  $\mu$ . But, since by assumption  $xy \in \lambda$ , we now have a cycle in  $C_{\lambda \wedge \mu}^+$  involving  $xy$ , which is again a contradiction given our assumptions.  $\square$

Overall, axioms R<sub>1</sub><sup>r</sup> and R<sub>2</sub><sup>r</sup> guarantee that  $\lambda \circ \mu$  always produces a result of the right format, uses only information present in  $\mu$  and  $\lambda$  to obtain it, and, by Proposition 2, preserves the acyclic part of  $\lambda \wedge \mu$ , which we may reasonably think of as uncontroversial and deserving to be withheld. Thus, through the last observation, axioms R<sub>1</sub><sup>r</sup> and R<sub>2</sub><sup>r</sup> recover another mainstay of AGM revision: if new information  $\mu$  is consistent with prior information  $\lambda$ , the result is, reassuringly,  $\lambda \wedge \mu \equiv \lambda$ , since in this case none of the comparisons of  $\lambda$  is involved in a cycle with  $\mu$ .

Axiom R<sub>3</sub><sup>r</sup> expresses the familiar notion of irrelevance of syntax. Axiom R<sub>4</sub><sup>r</sup> can be thought of as a compressed version of axioms R<sub>5</sub><sup>i</sup> and R<sub>6</sub><sup>i</sup> and expresses the idea that if  $\mu_2$  does not contradict  $\lambda \circ \mu_1$ , then revising by  $\mu_1$  and adding  $\mu_2$  is equivalent to revising by  $(\mu_1 \wedge \mu_2)$ .

Having presented the axioms, we want to move on now to a preference-driven mechanism for performing preference

revision—a mechanism that uses, as advertised, preferences over the adjacent comparisons of  $\lambda$ . We do this through a *resolute preference assignment*, i.e., a function  $f: \mathcal{C}_A \rightarrow \mathcal{T}_A$  mapping every  $\lambda \in \mathcal{C}_A$  to a strict linear order  $<_\lambda$  on the set  $A_\lambda$  of adjacent comparisons of  $\lambda$ . We further require  $<_\lambda$  to be insensitive to syntax, i.e., to satisfy property  $f_3^i$  adapted in the obvious way to the present context, in which case we call the assignment *faithful*. Intuitively, a faithful ranking  $<_\lambda$  requires the adjacent comparisons of  $\lambda$  to be ranked in a linear fashion, with the intention of formalizing a revision mechanism that uses this ranking in order to iteratively construct the new preference order: we read a statement like  $x_i x_{i+1} <_\lambda x_j x_{j+1}$  as saying that  $x_i x_{i+1}$  is more intensely held, or less eagerly given up, than  $x_j x_{j+1}$ . We use preferences on adjacent comparisons under the assumption that they are basic in the sense that they cannot be inferred by transitivity using other comparisons in  $\lambda$ , and therefore likely to be the result of explicit information.

Given a linear order  $<_\lambda$ , level  $i$  of  $A_\lambda$  according to  $<_\lambda$ , denoted  $\text{lev}_i(<_\lambda)$ , is defined as follows:

$$\begin{aligned} \text{lev}_{<_\lambda}^1(A_\lambda) &= \min_{<_\lambda}(A_\lambda), \\ \text{lev}_{<_\lambda}^{i+1}(A_\lambda) &= \min_{<_\lambda} \left( A_\lambda \setminus \left( \bigcup_{1 \leq j \leq i} \text{lev}_{<_\lambda}^j(A_\lambda) \right) \right). \end{aligned}$$

Intuitively, level  $i$  contains the  $i^{\text{th}}$  best comparison on  $A$  according to  $<_\lambda$ . Note that the levels of  $<_\lambda$  partition  $A_\lambda$  and, since  $A_\lambda$  is finite, there exists an index  $j > 0$  such that  $\text{lev}_{<_\lambda}^i(A_\lambda) = \emptyset$ , for all  $i \geq j$ . The *addition operator*  $\text{add}_{<_\lambda}^i(\mu)$  is defined as follows:<sup>2</sup>

$$\begin{aligned} \text{add}_{<_\lambda}^0(\mu) &= (C_\mu \cup CF_\mu(\lambda))^+, \\ \text{add}_{<_\lambda}^i(\mu) &= \begin{cases} (\text{add}_{<_\lambda}^{i-1}(\mu) \cup (\text{lev}_{<_\lambda}^i(A_\lambda)))^+, & \text{if acyclic,} \\ \text{add}_{<_\lambda}^{i-1}(\mu), & \text{otherwise.} \end{cases} \end{aligned}$$

Intuitively,  $\text{add}$  starts with the comparisons of  $\mu$ , plus the cycle-free part of  $\lambda$  with respect to  $\mu$ , and, at every further step  $i > 0$ , tries to add the comparison on level  $i$ : if the resulting set of comparisons does not contain a cycle (by taking its transitive closure) the operation is successful, and the new comparison is added; if not, the addition operator does nothing. Since the addition of new comparisons follows the order  $<_\lambda$ , this ensures that more highly valued comparisons are considered before lower quality ones. Note that  $\text{add}_{<_\lambda}^0(\mu) \subseteq \text{add}_{<_\lambda}^1(\mu) \subseteq \dots$ . Also note that the number of non-empty levels of  $A_\lambda$  is finite and the addition operation eventually reaches a fixed point, i.e., there exists  $j \geq 0$  such that  $\text{add}_{<_\lambda}^i(\mu) = \text{add}_{<_\lambda}^j(\mu)$ , for any  $i \geq j$ . We denote by  $\text{add}_{<_\lambda}^*(\mu)$  the fixed point of this operator and first check that it actually results a linear order.

**Proposition 3.** *If  $\lambda \in \mathcal{C}_A$ ,  $\mu \in \mathcal{P}_A$ , and  $<_\lambda$  is a faithful linear order on  $A_\lambda$ , then  $\text{add}_{<_\lambda}^*(\mu)$  is a strict linear order.*

*Proof.* The relation  $\text{add}_{<_\lambda}^*(\mu)$  is irreflexive and transitive by construction. Consider, next, two distinct alternatives  $x$  and

<sup>2</sup>The addition operator is different from the eponymous operation in (Benferhat et al. 1993).

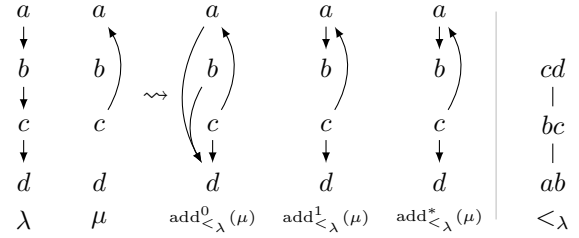


Figure 2: Revision according to a faithful linear order  $<_\lambda$ . Lower comparisons are better. Comparisons inferred by transitivity are omitted.

$y$ . If  $xy$  (or  $yx$ ) is in  $C_\mu^+$  or  $CF_\mu(\lambda)$  then  $xy$  (or  $yx$ ) is in  $\text{add}_{<_\lambda}^0(\mu) \subseteq \text{add}_{<_\lambda}^*(\mu)$ . If neither  $xy$ , nor  $yx$ , is in either of  $C_\mu^+$  or in  $CF_\mu(\lambda)$ , then we can assume wlog that  $xy$  is implied by  $\lambda$  (either  $xy$  or  $yx$  has to be implied by  $\lambda$ , since  $\lambda$  describes a linear order), and thus there exist  $t \geq 0$  adjacent comparisons  $z_0 z_1, \dots, z_{t-1} z_t$  in  $\lambda$ ,  $z_0 = x$  and  $z_t = y$ , that imply  $xy$  by transitivity. By our present assumption,  $z_0 z_1, \dots, z_{t-1} z_t$  are involved in a cycle with the edges of  $\mu$ . Since they are linearly ordered in  $<_\lambda$ , the addition operator goes through them successively:  $t - 1$  out of them eventually get chosen, and thus either  $xy$  or  $yx$  ends up inferred at some point. In all cases, though, either  $xy$  or  $yx$  is in  $\text{add}_{<_\lambda}^*(\mu)$ , which shows that  $\text{add}_{<_\lambda}^*(\mu)$  is connected.  $\square$

The proof of Proposition 3 shows that  $\text{add}_{<_\lambda}^*(\mu)$  is not just a linear order, but that it is obtained using only comparisons from  $\mu$  and  $\lambda$ , i.e., that it is a  $\lambda$ -model of  $\mu$ . This observation will come in handy later.

**Corollary 1.** *If  $\lambda \in \mathcal{C}_A$ ,  $\mu \in \mathcal{P}_A$ , and  $<_\lambda$  is a faithful linear order on  $A_\lambda$ , then  $\text{add}_{<_\lambda}^*(\mu) \in [\mu]_\lambda$ .*

For now, Proposition 3 gives us all we need to define a resolute preference revision operator: the *f-induced preference revision operator*  $\circ^f$  is defined as:

$$[\lambda \circ^f \mu] = \{\text{add}_{<_\lambda}^*(\mu)\}.$$

**Example 7.** *Take  $\lambda$  and  $\mu$  as in Example 5, with  $[\lambda] = \{abcd\}$  and  $C_\mu^+ = \{ca\}$ , and  $<_\lambda$  depicted in Figure 2. The order  $\lambda \circ \mu$  is assembled in steps, starting with the comparisons in  $\mu$  and the cyclic-free part of  $\lambda$ :  $C_\mu \cup CF_\mu(\lambda) = (\{ca\} \cup \{ad, bd, cd\})^+$ , depicted in Figure 2. Then, in the first step  $ab$  is added; the second step tries to add  $bc$ , but finds that this leads to a cycle with the previously added comparisons; the third step adds  $cd$ , which had been added already, and the process stops. The result is  $[\lambda \circ^f \mu] = \{cabd\}$ .*

The next step is to show that the constructive procedure using the addition operator fits within the framework delineated by axioms  $R_1^r$ – $R_4^r$ . This involves proving a representation result consisting of two parts: on the one hand, we show that an  $f$ -induced revision operator satisfies axioms  $R_1^r$ – $R_4^r$ ; in the second part, we show that a resolute operator assumed to satisfy axioms  $R_1^r$ – $R_4^r$  can be rationalized using a resolute faithful assignment. This involves reconstructing  $<_\lambda$  one pair of adjacent comparisons at a time, and involves

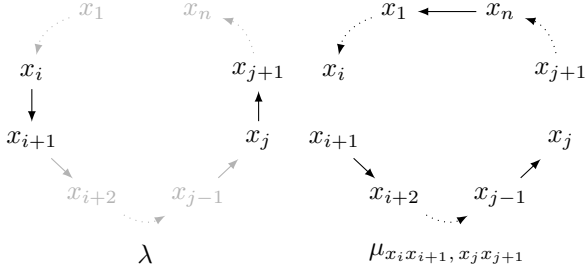


Figure 3: To force a choice between  $x_i x_{i+1}$  and  $x_j x_{j+1}$  we revise by  $\mu_{x_i x_{i+1}, x_j x_{j+1}}$ : the comparison that survives is assumed to be the more preferred.

solving an issue taken for granted in propositional revision: how to decide which is better of two adjacent comparisons? We address this issue by creating a special type of preference order that, when added to  $\lambda$ , forces a choice between any two adjacent comparisons.

Thus, if  $\ell = x_1 \dots x_n$  is the model of  $\lambda$  and  $x_i x_{i+1}$  and  $x_j x_{j+1}$  are adjacent comparisons in  $\ell$ , with  $i < j$ , the preference statement  $\mu_{x_i x_{i+1}, x_j x_{j+1}}$  is defined as:

$$\mu_{x_i x_{i+1}, x_j x_{j+1}} = \bigwedge_{i+1 \leq s \leq j-1} (x_s \succ x_{s+1}) \wedge \bigwedge_{j+1 \leq t \leq i-1} (x_t \succ x_{t+1}),$$

with the convention that  $x_s x_s$  is ignored, if it occurs, and that indices are computed modulo  $n$ , i.e.,  $x_{n+k}$  stands for  $x_k$ . Intuitively,  $\mu_{x_i x_{i+1}, x_j x_{j+1}}$  defines a partial order that includes the paths from  $x_{i+1}$  to  $x_j$  and from  $x_{j+1}$  to  $x_1$  in  $\lambda$  (see Figure 3). Under the assumption that revision has to create a  $\lambda$ -model of  $\mu$ , the gadget  $\mu_{x_i x_{i+1}, x_j x_{j+1}}$  forces us to make a choice between  $x_i x_{i+1}$  and  $x_j x_{j+1}$ .

**Proposition 4.** *If  $\circ$  is a resolute preference revision operator satisfying axioms  $R_1^r$  and  $R_2^r$ , then exactly one of  $x_i x_{i+1}$  and  $x_j x_{j+1}$  is in  $x_1 \in \lambda \circ \mu_{x_i x_{i+1}, x_j x_{j+1}}$ .*

*Proof.* At least one of  $x_i x_{i+1}$  and  $x_j x_{j+1}$  must be added, since they are adjacent comparisons and cannot be otherwise inferred from other comparisons through transitivity; but it is impossible to add both, as this would lead to a cycle.  $\square$

Thus, exactly one of the two comparisons ends up in  $\lambda \circ \mu_{x_i x_{i+1}, x_j x_{j+1}}$ : the one, we want to say, that is held more intensely. With this we can finally prove our main result.

**Theorem 3.** *A resolute preference revision operator  $\circ$  satisfies axioms  $R_1^r$ – $R_4^r$  iff there exists a resolute faithful assignment mapping every complete preference statement  $\lambda$  to a linear order  $<_\lambda$  on  $A_\lambda$  such that  $[\lambda \circ \mu] = \{\text{add}_{<_\lambda}^*(\mu)\}$ .*

*Proof.* For one direction we show that if  $f$  is a resolute faithful assignment, the  $f$ -induced operator  $\circ^f$  satisfies the axioms. Corollary 1 shows that  $\circ^f$  satisfies axioms  $R_1^r$  and  $R_2^r$ . Axiom  $R_3^r$  follows using the insensitivity to syntax of  $<_\lambda$ . For axiom  $R_4^r$  note that if  $(\lambda \circ \mu_1) \wedge \mu_2$  is consistent, then all the comparisons in  $\mu_2$  are in  $\lambda \circ \mu_1$ , since the latter encodes

a linear order. This means that all the comparisons in  $\lambda$  that get added to  $\mu_1$  to form  $\lambda \circ \mu_1$  can be added to  $\mu_2$  as well.

For the other direction take a resolute operator  $\circ$  satisfying all the postulates, and define  $<_\lambda$  on  $A_\lambda$  as follows:

$$x_i x_{i+1} <_\lambda x_j x_{j+1} \text{ if } x_i x_{i+1} \in \lambda \circ \mu_{x_i x_{i+1}, x_j x_{j+1}}.$$

Axioms  $R_1^r$  and  $R_2^r$  imply, via axioms  $R_1^r$ – $R_2^r$ , that  $<_\lambda$  is a strict, total order on  $A_\lambda$ . To show that  $<_\lambda$  is transitive, take three adjacent comparisons such that  $x_i x_{i+1} <_\lambda x_j x_{j+1} <_\lambda x_k x_{k+1}$ . From the facts that  $x_i x_{i+1} \in \lambda \circ \mu_{x_i x_{i+1}, x_j x_{j+1}}$  and that  $x_j x_{j+1} \in \lambda \circ \mu_{x_j x_{j+1}, x_k x_{k+1}}$  we infer, using axiom  $R_4^r$  that  $x_i x_{i+1} \in \mu_{x_i x_{i+1}, x_j x_{j+1}, x_k x_{k+1}}$ , where  $x_i x_{i+1} \in \mu_{x_i x_{i+1}, x_j x_{j+1}, x_k x_{k+1}}$  is a preference statement defined, as expected, as a cycle going from  $x_1$  to  $x_n$  to  $x_1$ , from which we remove comparisons  $x_i x_{i+1}$ ,  $x_j x_{j+1}$  and  $x_k x_{k+1}$ . Then, by adding  $x_j x_{j+1}$  to  $x_i x_{i+1} \in \mu_{x_i x_{i+1}, x_j x_{j+1}, x_k x_{k+1}}$ , we infer, again using axiom  $R_4^r$ , that  $x_i x_{i+1} \in \lambda \circ \mu_{x_i x_{i+1}, x_k x_{k+1}}$ .

For the final step, we have to show that  $[\lambda \circ \mu] = \{\text{add}_{<_\lambda}^*(\mu)\}$ . Assume, to the contrary, that there exists  $xy \in C_{\lambda \circ \mu}^+$  and that  $xy \notin \text{add}_{<_\lambda}^*(\mu)$ . This implies that  $yx \in \text{add}_{<_\lambda}^*(\mu)$ . We further infer that neither  $xy$  nor  $yx$  is in  $C_\mu^+$  (if  $yx \in C_\mu^+$ , then  $xy$  could not be in  $C_{\lambda \circ \mu}^+$ , which has to be, by axiom  $R_1^r$ , a  $\lambda$ -model of  $\mu$ ). Then either  $yx$  gets added by the addition operator directly, as an adjacent comparison of  $\lambda$ , or is inferred by transitivity. If it is added as an adjacent comparison of  $\lambda$ , we first note that it must be involved in a cycle with  $\mu$  (otherwise it would be in  $C_{\lambda \circ \mu}^+$ , per Proposition 2), and that it must have priority in  $<_\lambda$  over some other adjacent comparison  $zt$ , which must be sacrificed in order to break the cycle, i.e.,  $yx <_\lambda zt$ . But, from the fact that  $xy \in C_{\lambda \circ \mu}^+$  we can infer, using axiom  $R_4^r$ , that  $xy \in C_{\lambda \circ \mu_{y x, z t}}^+$ , which then implies that  $zt <_\lambda yx$ , a contradiction. If  $yx$  is inferred by transitivity rather than added directly, the reasoning is similar.  $\square$

## 5 Conclusions

We have presented two models of preference change. The first works by minimizing distances, and was found to be difficult to square with the requirement that prior and revised preference information are of the same type. The second, advanced in part to address this issue, is based on the idea that revising a preference  $\lambda$  goes hand in hand with having preferences over the comparisons inherent in  $\lambda$ . In formalizing these two approaches we believe to have provided a rigorous formal treatment to intuitions found elsewhere in the literature (Sen 1977; Grüne-Yanoff and Hansson 2009).

There is also ample space for future work, in particular with respect to understanding the distance functions underlying operators that are both axiom-abiding and that satisfy desirable constraints on the format of output. The initial work in Section 3 shows that combining these two types of requirements is a delicate matter, and further research is certain to yield interesting possibility or impossibility results. With respect to the framework of Section 4, an obvious way forward is to relax the assumption of linearity on the comparisons of  $\lambda$  in order to characterize choice mechanisms operating on a more general form of preference structure.

## Acknowledgments

We are grateful to the anonymous referees. This work was funded by NWO project OND1362710 and by the Austrian Science Fund (FWF): P30168.

## References

- Alchourrón, C. E.; Gärdenfors, P.; and Makinson, D. 1985. On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. *The Journal of Symbolic Logic*, 50(2): 510–530.
- Benferhat, S.; Cayrol, C.; Dubois, D.; Lang, J.; and Prade, H. 1993. Inconsistency management and prioritized syntax-based entailment. In *Proceedings of IJCAI 1993*, 640–645.
- Bentham, J.; and Liu, F. 2014. Deontic Logic and Preference Change. *IfCoLog Journal of Logics and their Applications*, 1(2): 1–46.
- Boutilier, C.; Brafman, R. I.; Domshlak, C.; Hoos, H. H.; and Poole, D. 2004. CP-nets: A Tool for Representing and Reasoning with Conditional Ceteris Paribus Preference Statements. *Journal of Artificial Intelligence Research (JAIR)*, 21: 135–191.
- Bradley, R. 2007. The kinematics of belief and desire. *Synthese*, 156(3): 513–535.
- Cadilhac, A.; Asher, N.; Lascarides, A.; and Benamara, F. 2015. Preference Change. *Journal of Logic, Language and Information*, 24(3): 267–288.
- Chomicki, J. 2003. Preference formulas in relational queries. *ACM Trans. Database Syst.*, 28(4): 427–466.
- Chomicki, J.; and Song, J. 2005. Monotonic and Nonmonotonic Preference Revision. In *Proc. IJCAI 2005 Multidisciplinary Workshop on Advances in Preference Handling*.
- Delgrande, J. P.; Peppas, P.; and Woltran, S. 2018. General Belief Revision. *Journal of the ACM (JACM)*, 65(5): 29:1–29:34.
- Dell’Acqua, P.; and Pereira, L. M. 2005. Preference Revision Via Declarative Debugging. In *Portuguese Conference on Artificial Intelligence*, 18–28. Springer.
- Delobelle, J.; Haret, A.; Konieczny, S.; Mailly, J.; Rossit, J.; and Woltran, S. 2016. Merging of Abstract Argumentation Frameworks. In *Proceedings KR 2016*, 33–42.
- Diller, M.; Haret, A.; Linsbichler, T.; Rümmele, S.; and Woltran, S. 2015. An Extension-Based Approach to Belief Revision in Abstract Argumentation. In *Proceedings of IJCAI 2015*, 2926–2932.
- Domshlak, C.; Hüllermeier, E.; Kaci, S.; and Prade, H. 2011. Preferences in AI: An Overview. *Artificial Intelligence*, 175(7-8): 1037–1052.
- Fermé, E. L.; and Hansson, S. O. 2018. *Belief Change: Introduction and Overview*. Springer Briefs in Intelligent Systems. Springer.
- Frankfurt, H. G. 1988. Freedom of the Will and the Concept of a Person. In *What is a person?*, 127–144. Springer.
- Freund, M. 2004. On the revision of preferences and rational inference processes. *Artificial Intelligence*, 152(1): 105–137.
- Grüne-Yanoff, T. 2013. Preference change and conservatism: comparing the Bayesian and the AGM models of preference revision. *Synthese*, 190(14): 2623–2641.
- Grüne-Yanoff, T.; and Hansson, S. O. 2009. From Belief Revision to Preference Change. In *Preference Change: Approaches from Philosophy, Economics and Psychology*, 159–184. Springer.
- Grüne-Yanoff, T.; and Hansson, S. O., eds. 2009. *Preference Change: Approaches from Philosophy, Economics and Psychology*, volume 42 of *Theory and Decision Library A*. Springer.
- Hansson, S. O. 1995. Changes in preference. *Theory and Decision*, 38(1): 1–28.
- Haret, A.; and Wallner, J. P. 2021. An AGM Approach to Revising Preferences. In *Proceedings of NMR 2021*, 41–50.
- Haret, A.; Wallner, J. P.; and Woltran, S. 2018. Two Sides of the Same Coin: Belief Revision and Enforcing Arguments. In *Proceedings IJCAI 2018*, 1854–1860.
- Harsanyi, J. C. 1955. Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *Journal of Political Economy*, 63(4): 309–321.
- Jeffrey, R. C. 1974. Preference among preferences. *Journal of Philosophy*, 71(13): 377–391.
- Katsuno, H.; and Mendelzon, A. O. 1992. Propositional Knowledge Base Revision and Minimal Change. *Artificial Intelligence*, 52(3): 263–294.
- Kendall, M.; and Gibbons, J. D. 1990. *Rank Correlation Methods*. New York: Oxford University Press.
- Lang, J.; and van der Torre, L. W. N. 2008. Preference Change Triggered by Belief Change: A Principled Approach. In *Proceedings of LOFT 8*, 86–111.
- Liu, F. 2011. *Reasoning About Preference Dynamics*, volume 354 of *Synthese Library*. Springer.
- Ma, J.; Benferhat, S.; and Liu, W. 2012. Revising Partial Pre-Orders with Partial Pre-Orders: A Unit-Based Revision Framework. In *Proceedings of KR 2012*, 633–637.
- Nebel, B. 1992. Syntax-Based Approaches to Belief Revision. In Gärdenfors, P., ed., *Belief Revision*, 52–88. Cambridge: Cambridge University Press.
- Nozick, R. 1994. *The Nature of Rationality*. Princeton University Press.
- Pigozzi, G.; Tsoukiàs, A.; and Viappiani, P. 2016. Preferences in artificial intelligence. *Annals of Mathematics and Artificial Intelligence*, 77(3-4): 361–401.
- Rossi, F.; Venable, K. B.; and Walsh, T. 2011. *A Short Introduction to Preferences: Between Artificial Intelligence and Social Choice*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers.
- Russell, S. 2019. *Human Compatible: Artificial Intelligence and the Problem of Control*. Penguin.
- Sen, A. K. 1977. Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. *Philosophy & Public Affairs*, 317–344.
- Sen, A. K. 2017. *Collective Choice and Social Welfare: Expanded Edition*. Penguin UK.