# The Compromise and Attraction Effects Through Frame Preferences

Alex Poterack*

November 13, 2015

**Abstract**

The compromise and attraction effects are two of the most robust and well documented violations of WARP, typically arising in the context of goods which can be judged along several distinct attributes. I construct a novel method of representing them by reducing the context of each menu to a "frame," representing the worst option along each attribute in the menu, and creating a collection of preferences indexed by frames. The preferences behave as though a good's attractiveness along each attribute is judged relative to the frame with declining marginal utility. I also characterize the properties of a function which represents this collection of preferences. Finally, I give a representation theorem characterizing the set of preferences represented by such a function.

## 1 Introduction

Among the well-documented violations of the standard choice axioms are effects of menu and context not included in canonical theory. In the standard model of decision making, agents are aware of the entire universe of goods, with a clear personal ranking of said goods. Such an agent should, therefore, be immune to any effects related to the inclusion or exclusion of a good from a menu. However, contra this standard model, there is mounting evidence[1] of such effects. Decision makers appear to judge alternatives relative to what *is* in front of them, rather than what *could be*.

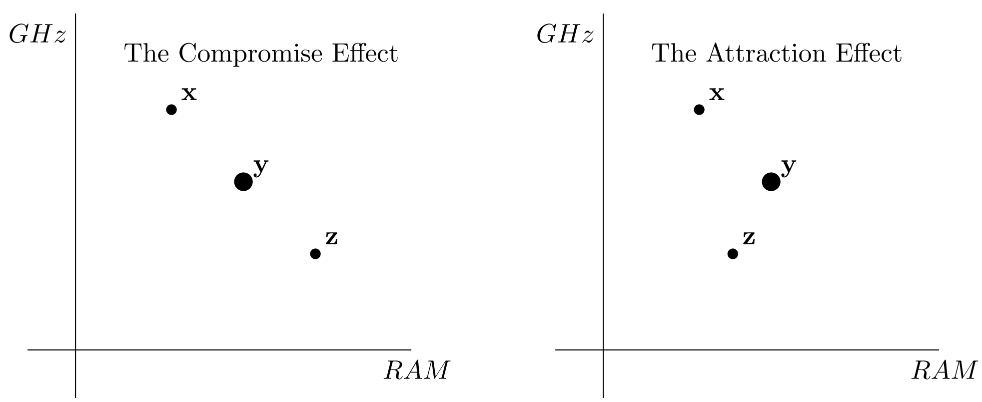
One circumstance in which such effects arise is when goods are comparable along several distinct attributes. For example, computers can be compared in terms of memory and processor speed; televisions can be compared in terms of size and picture quality; cars can be compared in terms of gas mileage and cargo space. It is easy to determine which computer has a faster processor, or which

*Department of Economics, Boston University, poterack@bu.edu. I am grateful for comments from Bart Lipman, Jawwad Noor, Dilip Mookherjee, Larry Epstein, Eddie Dekel, and various seminar participants.

[1] The effects studied in this paper were first identified by Huber et al. (1982) and Simonson (1989).

television is bigger, or which car gets better gas mileage. What is difficult, however, is determining how much of one attribute to trade off for improvement in another. This difficulty is the essence of the effects studied in this paper; experimentally, it is observed that decisionmakers are influenced in their decision of how to trade off between attributes by the presence of information the standard model considers extraneous.

The notable examples of this phenomenon considered herein are the compromise and attraction effects, which are best explained by example. Consider a decision maker planning to purchase a laptop, choosing between laptop $\mathbf{x}$, with a 2.5 GHz processor and 4 GBs of RAM; laptop $\mathbf{y}$, with a 2 GHz processor and 6 GBs of RAM; and laptop $\mathbf{z}$, with a 1.8 GHz processor and 8 GBs of RAM. The compromise effect is when agents choose $\mathbf{x}$ out of the menu $\{\mathbf{x}, \mathbf{y}\}$, and $\mathbf{y}$, but not $\mathbf{x}$, from the menu $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$. That is, the presence of laptop $\mathbf{z}$ switches their choice from $\mathbf{x}$ to $\mathbf{y}$, in violation of WARP, because the presence of $\mathbf{z}$ as a "more extreme" option makes $\mathbf{y}$ appear to be a desirable "compromise."



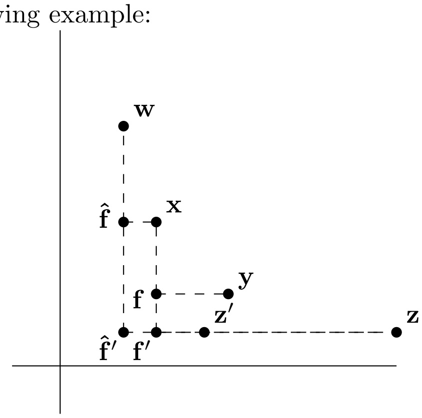

The "attraction effect" is similar. Suppose instead that laptop $\mathbf{z}$ represents a machine with a 1.8 GHz processor, but only 5 GBs of RAM. Some agents switch from choosing $\mathbf{x}$ out of $\{\mathbf{x}, \mathbf{y}\}$ to choosing $\mathbf{y}$ out of $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$. $\mathbf{z}$ is no longer "extreme" and making $\mathbf{y}$ a "compromise;" it is, however, clearly dominated on both attributes by $\mathbf{y}$ and only $\mathbf{y}$, thereby making $\mathbf{y}$ seem more "attractive" as an easy choice. Notice that in this example, the only difference between the effects is the horizontal position of $\mathbf{z}$.

In this example, decision makers must compare the marginal improvements of moving from a 2 GHz to a 2.5 GHz processor, or from 4 GBs of RAM to 6. The effects suggest that they draw inferences about how to judge these marginal improvements from the context created by the rest of the menu. One possible explanation for these phenomena is that this context is established by the worst good along each attribute.

Suppose the slowest processor speed establishes a "baseline" which all other processors are compared to, and the marginal utility of processing speed decreases moving further from the baseline. This is an idea introduced by Tversky and Kahneman (1991) in the context of loss aversion, which they call "diminishing sensitivity." In this case, the gain in moving from a 2 GHz processor to a 2.5 GHz processor would seem larger when the slowest processor is 2 GHz than it would in the presence of a 1.8 GHz processor. Intriguingly, both the compromise and attraction effects are consistent with this conceptualization.

This is the framework I will use. $\mathbb{R}^n$ is a set of goods[2]; the $n$ dimensions represent $n$ separate rankings over these goods. Each of these rankings is derived from an attribute along which goods are easily compared, such as memory or gas mileage. Decisionmakers make a choice from a subset of goods, which I will refer to as a "menu." There is a context created by this menu; following the example of Rubinstein and Salant (2008), I will call this context the "frame,"[3] and consider choice behavior which is standard for any fixed frame, but may demonstrate unusual effects when the frame is changed. More formally, there is a function mapping menus to frames, and when comparing choices from menus with the same frame, the decision makers' choices satisfy WARP. When comparing menus with different frames, WARP may be violated

The "baseline" interpretation illustrated in the laptop example naturally leads to a definition of the frame as the worst value for each attribute among goods in the menu (i.e., in the laptop example, it would be (1.8 GHz, 4 GB RAM) once laptop $\mathbf{z}$ is included). One attractive feature of this notion of frames is that the compromise and attraction effects both arise from the same source. Consider the following example:




Consider a choice correspondence $C(\cdot)$ such that $C(\{\mathbf{x}, \mathbf{y}\}) = \{\mathbf{x}, \mathbf{y}\}$. The compromise effect implies $C(\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) = \{\mathbf{y}\}$. The attraction effect implies $C(\{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}) = \{\mathbf{y}\}$. In both cases, the frame is lowered from $\mathbf{f}$ to $\mathbf{f}'$, and this change in frame is what changes the choice. The immediate observation is that lowering the frame makes $\mathbf{y}$ more appealing relative to $\mathbf{x}$.

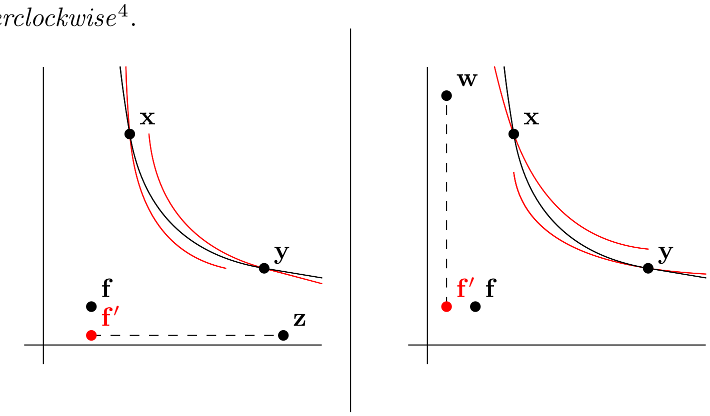
However, it would be a mistake to conclude lowering the frame makes $\mathbf{x}$ less appealing relative to *all* other goods. When adding $\mathbf{z}$ to the menu $\{\mathbf{w}, \mathbf{x}\}$, it is still the case that the frame is lowered; however, now this causes $\mathbf{x}$ to be chosen

[2]In the body of the paper, I consider the preferences over $\mathbb{R}^2$; I expand the model to consider $\mathbb{R}^n$ in Appendix A.

[3]This could also be called a "reference;" it fits naturally with the literature on reference dependence.

over **w**. In other words, compromise and attraction effects are consistent with a lowering of the frame making **x** *less* appealing relative to goods to its *right*, and *more* appealing relative to goods to its *left*.

This relationship is clearer when translated into terms of preferences. The assumption that WARP holds when the frame is held constant implies the existence of a collection of complete and transitive preferences indexed by frames. Denote this collection $\{\succsim^{\mathbf{f}}\}_{\mathbf{f}\in\mathbb{R}^2}$, where $\succsim^{\mathbf{f}}$ is the preference revealed by choices from menus with the frame $\mathbf{f}$. In the language of preferences, to say lowering the frame makes **x** less appealing relative to goods to its right and more appealing relative to goods to its left is to say it *rotates* the indifference curve associated with a given frame's revealed preference *clockwise* as the given frame is lowered. Similar analysis shows that moving the frame *left* rotates the curve *counterclockwise*[4].




Thus, this choice of frame definition and description of indifference curve rotation combine to form a succinct and intuitive description of the compromise and attraction effects. Furthermore, there is a straightforward mathematical interpretation. The collection of preferences $\{\succsim^{\mathbf{f}}\}_{\mathbf{f}\in\mathbb{R}^2}$ can be represented by a function of the form $U(\mathbf{x}, \mathbf{f})$. The rotation of indifference curves is a change in their slope. The slope of an indifference curve (holding the frame constant) is[5] $\frac{U_1}{U_2}$, and the desired rotation with regard to the frame is equivalent to:

$$\frac{\partial}{\partial 3}\frac{U_1}{U_2} > 0$$
$$\frac{\partial}{\partial 4}\frac{U_1}{U_2} < 0$$

It will be shown that this is equivalent to $\frac{U_{13}}{U_{23}} > \frac{U_1}{U_2} > \frac{U_{14}}{U_{24}}$, a property I will call *Compromise/Attraction Rotation*. I will show in the body of the paper this is closely related to the notion of "diminishing sensitivity."

[4]Consider repeating the analysis from the previous two paragraphs by adding **w** to the menu $\{\mathbf{x}, \mathbf{y}\}$.

[5]Where $U_i \equiv \frac{\partial U}{\partial i}$, i.e., the partial derivative of $U$ with respect to the $i$th argument.

The compromise and attraction effects only arise in choices from menus of three or more goods. As such, I will assume that the preference revealed by choices from menus containing only two goods, hereafter referred to as the "pairwise preference," is complete on $\mathbb{R}^2$, transitive, and continuous. In addition to limiting the departures from the standard model to the behavior of interest, this provides structure for the frame preferences revealed by larger menus. An inherent limitation of the frame preference notion is that each $\succsim^{\mathbf{f}}$ is complete only over a subset of $\mathbb{R}^2$; certain goods cannot be in a set with frame $\mathbf{f}$. It is difficult to impose regularity on how preferences change with respect to the frame when there are certain goods over which preferences are only defined for one frame and not the other. By generating a preference over all of $\mathbb{R}^2$ from the frame preferences, this assumption allows for desired regularity, namely monotonicity and continuity in the frame.

After a brief literature review in Section 2, the paper proceeds in Section 3.1 by taking the choice function and space of goods as a primitive, then deriving the frame preferences in Section 3.2. From there, Section 3.3 enumerates a list of properties for the utility function representation of the collection of frame preferences, followed by the equivalent behavioral axioms in Section 3.4. A representation theorem is proven in Section 3.5, followed by a discussion of the frame definition in Section 3.6.

## 2 Literature Review

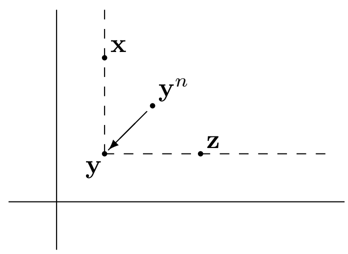
The attraction effect was first demonstrated experimentally by Huber et al. (1982), and the compromise effect was demonstrated first by Simonson (1989), whose paper also provided support for the attraction effect. These papers are strictly concerned with observing the effects; neither of them construct a representation incorporating these effects. An early model is Simonson and Tversky (1993), which (unlike mine) depends on context created by all elements in a menu, not just the worst along each attribute. Kivetz et al. (2004) analyze the Simonson and Tversky model, and two others, in terms of which best fit available experimental data. These papers propose several representations which demonstrate the effects; however, there is no axiomatization for any of these models, so their full behavioral implications remain unknown.

Ravid (2015) introduces a random choice procedure which allows for both effects. He shows this procedure can approximate a deterministic choice model if the deterministic model is equivalent to Simonson and Tversky (1993); it too depends on context created by all elements in a menu.

Ok et al. (2015) develop an axiomatized model which incorporates the attraction effect. Unlike my model, they endogenize the choice of reference (or frame). By construction, their model *cannot* represent the compromise effect. They have an axiom called "Reference Coherence" which essentially constructs a world where a good which "helps" another good in one context can never "harm" it in another context. However, the compromise effect helps a given good relative to some goods, but harms it relative to others. When $\mathbf{z}$ is down and to the right of **x**, goods to **x**'s right, between it and **z**, appear to be compromises, and **z** can make them preferred over **x**; i.e., **z** harms **x** relative to those goods. But **x** is between **z** and goods to the left of **x**, so compared to those goods, **x** is the compromise, and **z** can make **x** preferred over goods to its left; i.e., **z** helps **x** relative to those goods.

Barbos (2010) also gives a representation theorem admitting the attraction effect but not the compromise effect. In that paper, goods are divided into exogenous categories, and the attraction effect privileges goods which have an inferior good within their category. Again, there isn't a clear way to fit the compromise effect into this framework.

The first paper that axiomatizes a choice representation which incorporates both the attraction and compromise effects is de Clippel and Eliaz (2012). They construct a representation based around a multiple selves bargaining game, where the ranking of a good along each attribute represents its attractiveness to a given self. In this representation, if the menu contains a good which neither self views as the worst option, it will be selected. This violates continuity, as illustrated in the example pictured:




If there is a sequence of goods converging to good **y** as pictured, every good in the sequence will be the sole choice out of a menu with **x** and **z**, but **y** will not be chosen out of the menu $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$. Furthermore, they require indifference between any **x** and **z** as depicted (neither dominant on both attributes), so there are no circumstances where adding an irrelevant alternative may switch a choice entirely from exclusively choosing one good to exclusively choosing a different good. My model has continuity, and allows for such switches.

The paper most closely related to the present work is Tserenjigmid (2015; first draft 2014). While independently conceived and executed, both papers consider representations of choice behavior in which diminishing sensitivity generates both the compromise and attraction effects. The representations differ in that he characterizes the more specific functional form $g(u(x_1) - u(f_1)) + g(w(x_2) - w(f_2))$, with the frame of reference defined as it is in this paper, whereas I characterize a more general class of representations. Also, I show how to expand the set of goods considered from $\mathbb{R}^2$ to $\mathbb{R}^n$, an extension Tserenjigmid also notes is feasible for his representation.

The additively separable form of the representation is useful and tractable; it is employed by Poterack and Solow (2015) to study electoral politics. Tserenjigmid uses two axioms for this structure: the Thomsen condition and a translation invariance condition. In addition to generating the convenient functional form, these axioms also rule out some preference profiles that are potentially of interest. The Thomsen condition applies to all preferences revealed by choices, even those revealed by choices from two-good menus which cannot display either compromise or attraction effects. The translation invariance condition places restrictions on the magnitude of the effects as the menu is translated in good space. This poses difficulty for applications which wish to consider menu effects whose magnitudes vary in the good space. For example, in the earlier given description of a decision maker purchasing a laptop, the compromise and attraction effects seem very plausible. However, if this decision maker were instead considering various options for a network of servers costing hundreds of thousands of dollars, the decision would presumably be much more carefully considered and thus less influenced by these effects. Because the additively separable form allows for no interaction between the components of the frame, it is not well suited for modeling both kinds of decisions.

The general concept of choice influenced by a frame is related to a class of models explored by Rubinstein and Salant (2008). Rubinstein and Salant consider choice functions which take both the menu *and* a frame as arguments; despite my emphasis on frames, the choice correspondence in this paper only takes the menu as an argument, because the frame is a *function* of the menu.

# 3 The Model

## 3.1 Primitives

$\mathbb{R}^2$ represents a set of goods[6]. The two components of an element of $\mathbb{R}^2$ represent two "attributes" which are easily compared (e.g., RAM and processor speed).

The decision maker has a choice correspondence $C : \mathcal{S} \to \mathcal{S}$, where $\mathcal{S}$ is the collection of all proper compact[7] subsets of $\mathbb{R}^2$. For simplicity, assume that a good with a larger number in one component is preferred along the represented attribute. Mathematically, if $x_{-i} = y_{-i}$, then $\mathbf{x} \in C(\{\mathbf{x}, \mathbf{y}\})$ if and only if $x_i \geq y_i$. Define the attribute preferences $\succsim_1$ and $\succsim_2$ by $\mathbf{x} \succ_i \mathbf{y}$ if and only if $x_i \geq y_i$. These preferences are complete and transitive[8].

## 3.2 Frames, Revealed Preferences, and fWARP

The choice correspondence $C(\cdot)$ can be used to construct a collection of revealed frame preferences. First, the concept of a "frame" on a menu must be defined.

[6]The model can be extended to $\mathbb{R}^n$; details are found in appendix A.
[7]Compactness is used solely for the sake of a well-defined minimum.
[8]Monotonicity in the components can be relaxed; see appendix B.

Define *"frame of S"* ($\mathbf{f}(S)$) by[9]

$$\mathbf{f}(S) \equiv \left(\min_{\mathbf{x}\in S} x_1, \min_{\mathbf{x}\in S} x_2\right) \tag{1}$$

Define a revealed frame preference, $\succsim^{\mathbf{f}}$ by

$$\mathbf{x} \succsim^{\mathbf{f}} \mathbf{y} \Leftrightarrow \exists\ S \in \mathcal{S} \text{ s.t. } \mathbf{x}, \mathbf{y} \in S, \mathbf{f}(S) = \mathbf{f}, \text{ and } \mathbf{x} \in C(S) \tag{2}$$

Note that by the definition of $\mathbf{f}(S)$, $\succsim^{\mathbf{f}}$ can only be defined for $\mathbf{x}, \mathbf{y} \in A^{\mathbf{f}}$, where

$$A^{\mathbf{f}} \equiv \left\{\mathbf{x} \in \mathbb{R}^2 | x_1 \geq f_1, x_2 \geq f_2\right\} \text{ (where } \mathbf{f} = (f_1, f_2)).$$

There exists an $\succsim^{\mathbf{f}}$ for each $\mathbf{f} \in \mathbb{R}^2$. Collectively, they are $\left\{\succsim^{\mathbf{f}}\right\}_{\mathbf{f}\in\mathbb{R}^2}$.

Obviously, to study the compromise and attraction effects, WARP must be weakened. However, some regularity is still desired. It is given by the following axiom:

**Axiom 1** (The Weak Axiom of Revealed Frame Preference (fWARP))**.** If for some $S \in \mathcal{S}$ with $\mathbf{x}, \mathbf{y} \in S$, $\mathbf{x} \in C(S)$, then for any $S' \in \mathcal{S}$ with $\mathbf{x}, \mathbf{y} \in S'$ and $\mathbf{f}(S) = \mathbf{f}(S')$, $\mathbf{y} \in C(S')$ implies $\mathbf{x} \in C(S')$.

This axiom is identical to standard WARP, save for the inclusion of the additional condition "$\mathbf{f}(S) = \mathbf{f}(S')$." This reflects the underlying idea the model: within a frame, preferences behave as expected. When the frame changes, the preferences do as well.

fWARP implies that $\left\{\succsim^{\mathbf{f}}\right\}_{\mathbf{f}\in\mathbb{R}^2}$ represents $C(\cdot)$ in the sense that

$$C(S) = \left\{\mathbf{x} \in S | \mathbf{x} \succsim^{\mathbf{f}(S)} \mathbf{y}\ \forall\ \mathbf{y} \in S\right\} \tag{3}$$

In light of this, I work with the frame preferences henceforth.

### 3.3 Properties

We will begin by considering the properties desirable in such a function. Firstly, properties to ensure tractability of the utility function are desired. Specifically, it should be monotone in the first two arguments, and continuous in all four arguments. Given that goods are assumed to be increasing in desirability in their components, the function must be strictly increasing in the first two arguments. No monotonicity restriction is placed on the third and fourth arguments because how the utility varies with $\mathbf{f}$, holding $\mathbf{x}$ constant, has no behavioral implications and is therefore irrelevant to the paper.; the relevant issue is how the *relative* values of the utilities for various goods change. The behavior of the representation with regard to the frame is entirely a statement about the behavior of the *collection* of preferences: the change when moving from $\succsim^{\mathbf{f}}$ to $\succsim^{\mathbf{f}'}$.

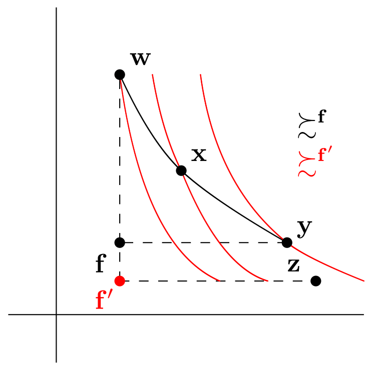
[9] This definition of frame is very specific, but it will be shown in Section 3.6 that it can be relaxed.

**Property 1** (*Regularity*)**.** $U$ is strictly increasing in the first two arguments and continuous in all four arguments.

Next is the property which gives the compromise and attraction effects, via the indifference curve rotation described in the introduction.

**Property 2** (*Compromise/Attraction Rotation*)**.** [10] $\frac{U_{13}}{U_{23}} > \frac{U_1}{U_2} > \frac{U_{14}}{U_{24}}$

This is the property which most captures the nature of the effects. It is equivalent to the indifference curves rotating clockwise when the frame is lowered, and counterclockwise as it is moved left. Considering the following example:




We would expect, by the compromise effect, the addition of $\mathbf{z}$ to the menu $\{\mathbf{w}, \mathbf{x}, \mathbf{y}\}$ to make $\mathbf{y}$ preferred to $\mathbf{x}$, but it would also make $\mathbf{x}$ preferred to $\mathbf{w}$. This corresponds to the clockwise rotation of indifference curves, in response to the frame being lowered from $\mathbf{f}$ to $\mathbf{f}'$. Similarly, the addition of $\mathbf{w}$ to the menu $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ would make $\mathbf{x}$ preferred to $\mathbf{y}$ and $\mathbf{y}$ preferred to $\mathbf{z}$. That corresponds to the counterclockwise rotation of indifference curves as the frame is moved left. Again, by the definition of frame used, this will also capture the attraction effect.

This is closely related to the property of "diminishing sensitivity," described by Tversky and Kahneman (1991). This property stipulates that the marginal value of a gain diminishes as it is further from the reference (or in this case, the frame). When the frame is lowered, the goods being compared are further from it. This means that a given advantage in the vertical attribute now has a lower value, and so a greater advantage is required to make up for a given disadvantage in the horizontal attribute—hence, the indifference curve must be steeper, which is achieved by its clockwise rotation. The same applies along the horizontal attribute when the frame is shifted left.

Finally, there are two more technical properties used for mathematical convenience. One guarantees transitivity among comparisons of pairs of elements; the other ensures that indifference curves cannot asymptote.

[10] The construction of this property suggests difficulty when derivatives are equal to zero or non-existent; this property can be expressed more generally to encompass these cases. Details can be found in the proof of the representation theorem.

**Property 3** (*Pairwise Transitivity*). $U(x_1, x_2, x_1, y_2) = U(y_1, y_2, x_1, y_2)$
& $U(y_1, y_2, y_1, z_2) = U(z_1, z_2, y_1, z_2)$
$\Rightarrow U(x_1, x_2, x_1, z_2) = U(z_1, z_2, x_1, z_2)$[11]

**Property 4** (*Non-asymptotic Indifference Curves*). Given a frame $\mathbf{f}$, for each $\mathbf{x} \in A^{\mathbf{f}}$ there exists a $\mathbf{y} \in A^{\mathbf{f}}$ such that

$$U(x_1, x_2, f_1, f_2) = U(f_1, y_2, f_1, f_2) = U(y_1, f_2, f_1, f_2)$$

## 3.4 Axioms

The crux of the idea is that preferences behave as usual when holding the frame fixed. As such, the first axiom says exactly that.

**Axiom 1** (*Continuous Weak Order*). $\succsim^{\mathbf{f}}$ is complete on $A^{\mathbf{f}}$, transitive, and continuous, $\forall\ \mathbf{f} \in \mathbb{R}^2$.

There are two more axioms which apply within a fixed frame, and do not consider moving the frame.

**Axiom 2** (*Simplicity*). $\mathbf{x} \succsim_1 \mathbf{y}$ and $\mathbf{x} \succsim_2 \mathbf{y} \Rightarrow \mathbf{x} \succsim^{\mathbf{f}} \mathbf{y}$ for all $\mathbf{f}$ such that $\mathbf{x}, \mathbf{y} \in A^{\mathbf{f}}$.

This is meant to capture the idea that choice is only difficult when the two attribute preferences disagree. If $\mathbf{x}$ is preferred to $\mathbf{y}$ on both attributes, it is always preferred to $\mathbf{y}$, regardless of context. It is equivalent to $U$ being strictly increasing in the first two arguments.

**Axiom 3** (*Substitutability*). Given $\mathbf{y}, \mathbf{f}$ such that $\mathbf{y} \in A^{\mathbf{f}}$, $\exists\ \mathbf{x}$ s.t. $x_1 = f_1$ and $\mathbf{y} \sim^{\mathbf{f}} \mathbf{x}$, and $\mathbf{z}$ s.t. $z_2 = f_2$ and $\mathbf{z} \sim^{\mathbf{f}} \mathbf{y}$.

This merely prevents asymptotic indifference curves; it is included for technical reasons, though it also captures the intuition that it's always possible to trade off between attributes. It is equivalent to Property 4.

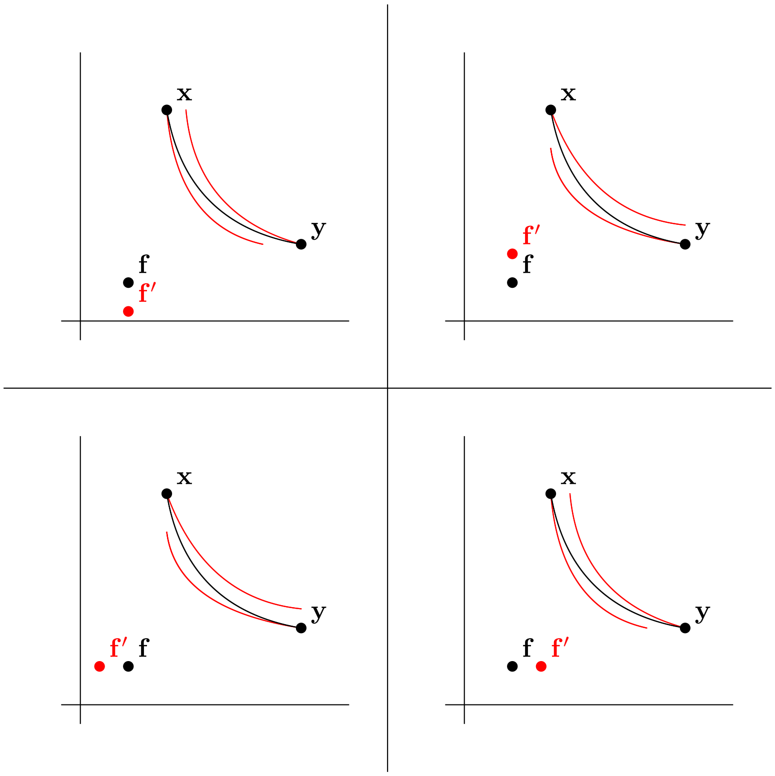
The next axiom is the first to address how preferences change when the frame is moved. It is equivalent to Property 2; as such, it is my formal statement of the compromise and attraction effects:

**Axiom 4** (*Compromise/Attraction Monotonicity*). Given $\mathbf{x} \sim^{\mathbf{f}} \mathbf{y}$ and $\mathbf{x} \succ_i \mathbf{y}$, then

1. $\mathbf{y} \succ^{(f_{-i}, f_i')} \mathbf{x}\ \forall\ f_i' < f_i$,
2. $\mathbf{x} \succ^{(f_{-i}, f_i'')} \mathbf{y}\ \forall\ f_i'' \in (f_i, y_i]$,
3. $\mathbf{x} \succ^{(f_{-i}', f_i)} \mathbf{y}\ \forall\ f_{-i}' < f_{-i}$, and

[11] This property exists to make Axiom 6, *Pairwise Weak Order*, necessary. If removed, *Pairwise Weak Order* is no longer necessary, but the rest of the axioms remain necessary, and *Pairwise Weak Order* remains sufficient.

4. $\mathbf{y} \succ^{(f''_{-i}, f_i)} \mathbf{x}\ \forall f''_{-i} \in (f_{-i}, x_{-i}]$.




A good which only lowers the frame cannot make $\mathbf{x}$ appear as a compromise relative to $\mathbf{y}$, nor be in position for the attraction effect to hold. Therefore, lowering the frame advantages $\mathbf{y}$. Similarly, a good which only moves the frame left cannot make $\mathbf{y}$ appear as a compromise relative to $\mathbf{x}$, nor be in position for the attraction effect to hold. Therefore, moving the frame left advantages $\mathbf{x}$.

The next axiom is a continuity axiom related to the behavior of the preferences as the frame changes. The same intuition which makes continuous preferences appealing also makes a continuity in how the preferences change with respect to the frame appealing. A small change in the frame should not make a sudden jump in the preferences. This intuition is captured by the following axiom:

**Axiom 5** (*Frame Continuity*)**.** Given $\mathbf{f}$, $\mathbf{x}, \mathbf{y}$ such that $\mathbf{x} \succ^{\mathbf{f}} \mathbf{y}$, and $f'_i$ such that $\mathbf{y} \succ^{(f_{-i}, f'_i)} \mathbf{x}$, then $\exists\ f''_i$ such that $\mathbf{x} \sim^{(f_{-i}, f''_i)} \mathbf{y}$.

As continuous preferences allow for continuous utility functions, *Frame Continuity* is a necessary condition for the representation to be continuous in $\mathbf{f}$. However, it is not a sufficient condition to guarantee continuity in $\mathbf{f}$, because each frame preference is incomplete on $\mathbb{R}^2$, and this creates issues with the behavior of the function as $\mathbf{f}$ changes. Given $\mathbf{f} \neq \mathbf{f}'$, $A^{\mathbf{f}} \neq A^{\mathbf{f}'}$, and the preferences over goods contained in one set but not the other may behave strangely. Axioms 4 and 5 restrict how preferences may *change* when the frame moves, but they do not affect goods in $(A^{\mathbf{f}} \cup A^{\mathbf{f}'}) \setminus (A^{\mathbf{f}} \cap A^{\mathbf{f}'})$, because these goods do not see preferences over them change; they see preferences over them being newly revealed. To put order on this process of preference revelation, I use an axiom which imposes the earlier stated desire to treat choices from menus of two goods as complete, transitive, and continuous – *Pairwise Weak Order*. This axiom requires the "pairwise revealed preference" $\succsim^*$, defined by the following:

$$\mathbf{x} \succsim^* \mathbf{y} \Leftrightarrow \mathbf{x} \in C(\{\mathbf{x}, \mathbf{y}\})$$

Note that $\mathbf{x} \succsim^* \mathbf{y} \Leftrightarrow \mathbf{x} \succsim^{\mathbf{f}(\{\mathbf{x},\mathbf{y}\})} \mathbf{y}$, so the completeness of the frame preferences implies the completeness of $\succsim^*$.

**Axiom 6** (*Pairwise Weak Order*)**.** $\succsim^*$ is transitive and continuous.

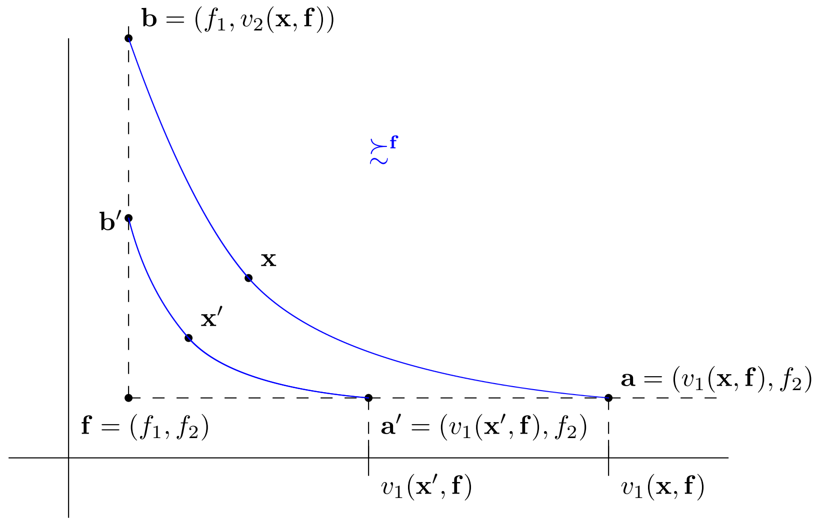
This creates a preference which applies to the whole space, yet also relates the frames to one another, providing more structure to regulate the behavior of the preferences when moving the frame. Along with *Frame Continuity*, *Pairwise Weak Order* is sufficient to guarantee continuity in $\mathbf{f}$, though it is not necessary for continuity in $\mathbf{f}$; this axiom also implies Property 3.

## 3.5 Representation Theorem

**Theorem 1** (Compromise/Attraction Representation Theorem)**.** Given a collection of preferences $\{\succsim^{\mathbf{f}}\}_{\mathbf{f} \in \mathbb{R}^2}$, there exists a function $U(\mathbf{x}, \mathbf{f})$ with Properties 1-4 representing it if and only if Axioms 1-6 hold.

Finding a utility function conditional on the frame is trivial; in fact, the first axiom alone guarantees the existence of one. Given a frame $\mathbf{f}$, by *Continuous Weak Order*, there exists $u^{\mathbf{f}}(\mathbf{x})$ which represents the $\mathbf{f}$-preference over $A^{\mathbf{f}}$. Repeat this for each $\mathbf{f} \in \mathbb{R}^2$, and define $U(\mathbf{x}, \mathbf{f}) = u^{\mathbf{f}}(\mathbf{x})$. Of course, this axiom alone does not deliver the other desired properties, namely continuity in $\mathbf{f}$, and therein lies the difficulty.

*Proof.* The proof first proposes a representation, then establishes some monotonicity and continuity properties of the representation in $\mathbf{f}$. Given a good $\mathbf{x}$ and a frame $\mathbf{f}$ such that $\mathbf{x} \in A^{\mathbf{f}}$, by *Substitutability*, there exists a good $\mathbf{a}(\mathbf{x})$ whose vertical position is $f_2$ such that $\mathbf{x} \sim^{\mathbf{f}} \mathbf{a}$. By *Simplicity* and transitivity of $\succsim^{\mathbf{f}}$, this $\mathbf{a}$ is unique. Define a function $v_1(\mathbf{x}, \mathbf{f})$ by $\mathbf{x} \sim^{\mathbf{f}} (v_1(\mathbf{x}, \mathbf{f}), f_2)$. $v_1(\mathbf{x}, \mathbf{f})$ is a representation of $\{\succsim^{\mathbf{f}}\}_{\mathbf{f} \in \mathbb{R}^2}$. To see this, consider an $\mathbf{x}'$ such that $\mathbf{x} \succ^{\mathbf{f}} \mathbf{x}'$, $\mathbf{x} \sim^{\mathbf{f}} (v_1(\mathbf{x}, \mathbf{f}), f_2)$ and $\mathbf{x}' \sim^{\mathbf{f}} (v_1(\mathbf{x}', \mathbf{f}), f_2)$, so by transitivity of $\succsim^{\mathbf{f}}$, $(v_1(\mathbf{x}, \mathbf{f}), f_2) \succ^{\mathbf{f}} (v_1(\mathbf{x}', \mathbf{f}), f_2)$. *Simplicity* then implies $v_1(\mathbf{x}, \mathbf{f}) > v_1(\mathbf{x}', \mathbf{f})$.




By *Simplicity*, $x_i > x_i'$ and $x_{-i} = x_{-i}'$ implies $\mathbf{x} \succ^{\mathbf{f}} \mathbf{x}'$ for each $\mathbf{f}$ such that $\mathbf{x}, \mathbf{x}' \in A^{\mathbf{f}}$. Therefore, $v_1$ is strictly increasing in $x_1$ and $x_2$.

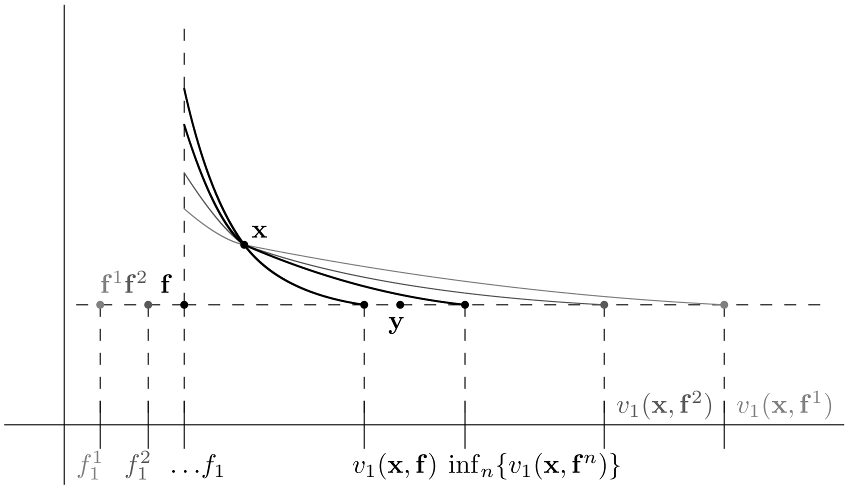
By continuity of $\succsim^{\mathbf{f}}$, $v_1$ is continuous in $x_1$ and $x_2$. To see why, consider $\{\mathbf{x}^n\} \to \mathbf{x}$. Without loss of generality, suppose $\{\mathbf{x}^n\}$ is weakly decreasing in both components. $v_1(\mathbf{x}, \mathbf{f}) \leq v_1(\mathbf{x}^n, \mathbf{f})$, for each $n$. Therefore,$\inf_n\{v_1(\mathbf{x}^n, \mathbf{f})\} \geq v_1(\mathbf{x}, \mathbf{f})$. If the inequality is strict, there exists $y_1 \in (v_1(\mathbf{x}, \mathbf{f}), \inf_n\{v_1(\mathbf{x}^n, \mathbf{f})\})$. $\{\mathbf{x}^n\}$ is in the upper contour set of $(y_1, f_2)$ for each $n$, but $\mathbf{x}$ is not in this upper contour set. This violates *Continuous Weak Order*, and therefore it must be the case that $\inf_n\{v_1(\mathbf{x}^n, \mathbf{f})\} = v_1(\mathbf{x}, \mathbf{f})$. This implies $v_1(\mathbf{x}^n, \mathbf{f}) \to v_1(\mathbf{x}, \mathbf{f})$, and thus the function is continuous in $x_1$ and $x_2$.

Importantly, $v_1$ is also continuous in $f_1$.

**Lemma 1.** Given axioms 1-5, $v_1$ is continuous in $f_1$.

*Proof.* Consider $\{f_1^n\}_{n=1}^{\infty} \to f_1$. Without loss of generality, suppose $\{f_1^n\}_{n=1}^{\infty}$ is increasing[12]. $v_1$ is decreasing in $f_1$. By *Compromise/Attraction Monotonicity*, moving the frame right rotates the indifference curves clockwise. Therefore, given $f_1' > f_1$, $\mathbf{f} \equiv (f_1, f_2)$, and $\mathbf{f}' \equiv (f_1', f_2)$, $(v_1(\mathbf{x}, \mathbf{f}), f_2) \succ^{\mathbf{f}'} (v_1(\mathbf{x}, \mathbf{f}'), f_2)$. This implies $v_1(\mathbf{x}, \mathbf{f}) > v_1(\mathbf{x}, \mathbf{f}')$, which implies $v_1$ is decreasing in $f_1$. Because $\{f_1^n\}_{n=1}^{\infty}$ is increasing, $v_1(\mathbf{x}, \mathbf{f}) \leq v_1(\mathbf{x}, \mathbf{f}^n)\ \forall\ n \in \mathbb{N}$. As $\{v_1(\mathbf{x}, \mathbf{f}^n)\}$ is bounded from below, there exists $\inf_n\{v_1(\mathbf{x}, \mathbf{f}^n)\}$, and $v_1(\mathbf{x}, \mathbf{f}) \leq \inf_n\{v_1(\mathbf{x}, \mathbf{f}^n)\}$.

[12]Because $\{f_1^n\}_{n=1}^{\infty} \to f_1$ is a sequence in $\mathbb{R}$, there exists a monotone subsequence which converges to $f_1$.




Suppose $v_1(\mathbf{x},\mathbf{f}) < \inf_n\{v_1(\mathbf{x},\mathbf{f}^n)\}$, as illustrated in the above picture. There exists a $\mathbf{y}$ such that $y_2 = f_2$ and $y_1 \in (v_1(\mathbf{x},\mathbf{f}), \inf_n\{v_1(\mathbf{x},\mathbf{f}^n)\})$. $\mathbf{y} \succ^{\mathbf{f}} \mathbf{x}$, but $\mathbf{x} \succ^{(f_1',f_2)} \mathbf{y} \;\forall\; f_1' < f_1$. Furthermore, by *Compromise/Attraction Monotonicity*, $\mathbf{y} \succ^{\mathbf{f}} \mathbf{x} \;\forall\; f_1' > f_1$. Therefore, there is no $\hat{f}_1$ such that $\mathbf{x} \sim^{(\hat{f}_1,f_2)} \mathbf{y}$, and this violates *Frame Continuity*. Thus, $v_1(\mathbf{x},\mathbf{f}) = \inf\{v_1(\mathbf{x},\mathbf{f}^n)\}$, it is the limit of $\{v_1(\mathbf{x},\mathbf{f}^n)\}$, and this implies $v_1$ is continuous in $f_1$. $\square$

Symmetric reasoning shows that there is also a function $v_2(\mathbf{x},\mathbf{f})$ defined by $\mathbf{x} \sim^{\mathbf{f}} (f_1, v_2(\mathbf{x},\mathbf{f}))$. $v_2$ is strictly increasing and continuous in $x_1$ and $x_2$, and continuous in $f_2$. So now there are two representations, $v_1$ and $v_2$. $v_1$ has the desired continuity in $f_1$, and $v_2$ has the desired continuity in $f_2$. The next step is to show these functions have the desired continuity in both arguments.

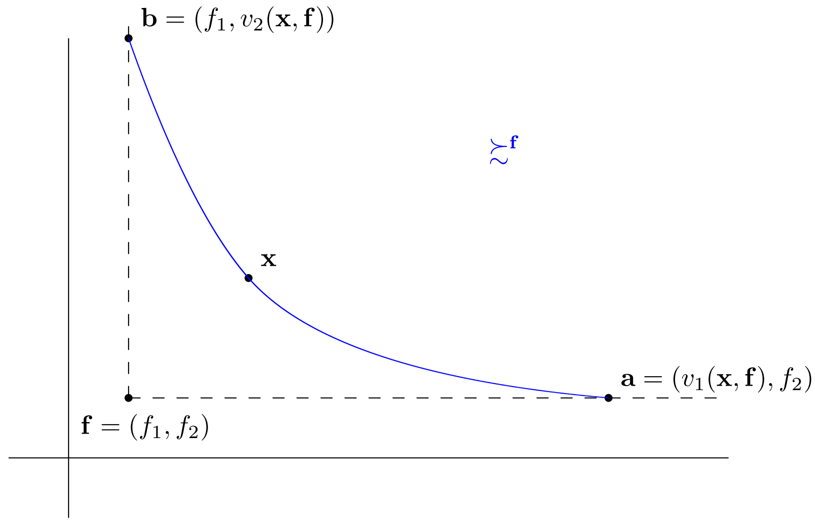
**Lemma 2.** Given $v_1$ as defined above, which is continuous in $f_1$, and $v_2$ as defined above, which is continuous in $f_2$, *Pairwise Weak Order* implies $v_1$ and $v_2$ are both continuous in both $f_1$ and $f_2$.

*Proof.* Consider the points $(v_1(\mathbf{x},\mathbf{f}), f_2) \equiv \mathbf{a}$ and $(f_1, v_2(\mathbf{x},\mathbf{f})) \equiv \mathbf{b}$. Both of these points are $\mathbf{f}$-indifferent to $\mathbf{x}$, which, by transitivity of $\succsim^{\mathbf{f}}$, means $\mathbf{a} \sim^{\mathbf{f}} \mathbf{b}$. Because[13] $\mathbf{f}(\{\mathbf{a},\mathbf{b}\}) = \mathbf{f}$, this means that $C(\{\mathbf{a},\mathbf{b}\}) = \{\mathbf{a},\mathbf{b}\}$, and thus the two points are pairwise indifferent, $\mathbf{a} \sim^* \mathbf{b}$.

$\succsim^*$ is continuous and transitive, and is complete by completeness of the $\mathbf{f}$-preferences. Therefore, there exists a continuous function $u^*(\cdot,\cdot)$ which represents the pairwise preference. Thus,

$$u^*(\mathbf{a}) = u^*(\mathbf{b})$$
$$u^*(v_1(\mathbf{x},\mathbf{f}), f_2) = u^*(f_1, v_2(\mathbf{x},\mathbf{f}))$$

[13] This equality is essential to the proof of this lemma, and will be important in Section 3.6, when other frame definitions are discussed.

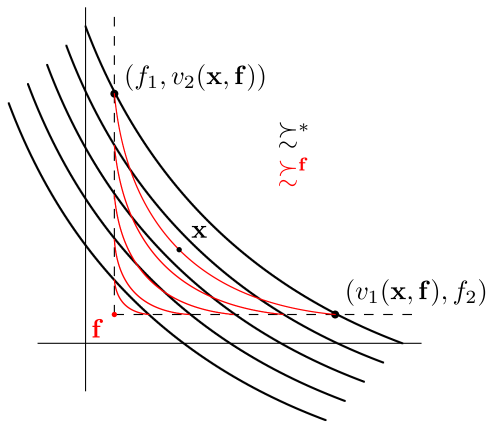



$u^*(\cdot,\cdot)$ can be used to demonstrate that $v_1$ is continuous in $f_2$. Consider $\{f_2^n\}_{n=1}^{\infty} \to f_2$. $v_2$ is continuous in $f_2$, so $\{v_2(\mathbf{x}, \mathbf{f}^n)\}_{n=1}^{\infty} \to v_2(\mathbf{x}, \mathbf{f})$ (where $\mathbf{f}^n \equiv (f_1, f_2^n)$). Therefore, continuity of $u^*$ implies $\{u^*(f_1, v_2(\mathbf{x}, \mathbf{f}^n))\}_{n=1}^{\infty} \to u^*(f_1, v_2(\mathbf{x}, \mathbf{f}))$. For each $n$, $u^*(v_1(\mathbf{x}, \mathbf{f}^n), f_2^n) = u^*(f_1, v_2(\mathbf{x}, \mathbf{f}^n))$, and $u^*(f_1, v_2(\mathbf{x}, \mathbf{f})) = u^*(v_1(\mathbf{x}, \mathbf{f}), f_2)$. This implies that $\{u^*(v_1(\mathbf{x}, \mathbf{f}^n), f_2^n)\}_{n=1}^{\infty} \to u^*(v_1(\mathbf{x}, \mathbf{f}), f_2)$. Furthermore, by *Simplicity*, $u^*(\cdot,\cdot)$ is strictly increasing in both arguments. Thus, in addition to being continuous in $f_1$, $v_1$ is continuous in $f_2$. The same proof can show that $v_2$ is continuous in $f_1$. □

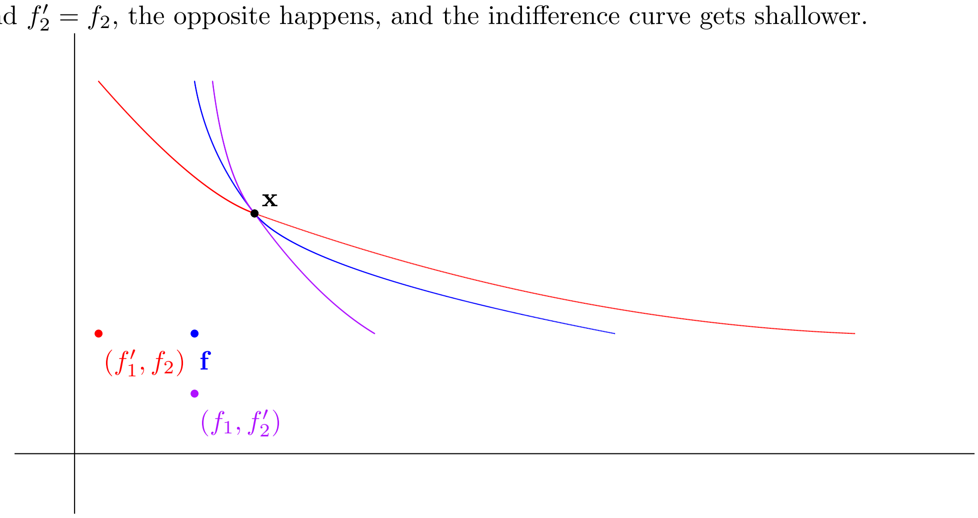
There are two functions which satisfy the continuity property required; they each define the function $U(\mathbf{x}, \mathbf{f})$ by

$$
\begin{aligned}
U(\mathbf{x}, \mathbf{f}) &= u^*(v_1(\mathbf{x}, \mathbf{f}), f_2) \\
&= u^*(f_1, v_2(\mathbf{x}, \mathbf{f}))
\end{aligned}
$$

In other words, the weak order of the pairwise preference “ties together” the $v_1$ and $v_2$ functions.




The preceding establishes that $U$ satisfies Property 1. $U$ satisfies Property 2 because this property is equivalent to *Compromise/Attraction Monotonicity*. Take a frame $\mathbf{f}$ and a point $\mathbf{x} \in A^{\mathbf{f}}$. Consider the points "left" of $\mathbf{x}$ (i.e., those points with smaller first components) which $\mathbf{x}$ is $\mathbf{f}$-indifferent to. If $\mathbf{f}$ is lowered to some $\mathbf{f}'$ with $f_2' < f_2$ and $f_1' = f_1$, by *Compromise/Attraction Monotonicity*, $\mathbf{x}$ is $\mathbf{f}'$-*preferred* to these points above it. Now consider the points "right" of $\mathbf{x}$ (with a larger first component) which $\mathbf{x}$ is $\mathbf{f}$-indifferent to. $\mathbf{x}$ is $\mathbf{f}'$-*dispreferred* to these points. In other words, the indifference curve passing through $\mathbf{x}$ gets steeper with the move to $\mathbf{f}'$. If instead $\mathbf{f}$ is moved left to some $\mathbf{f}'$ with $f_1' < f_1$ and $f_2' = f_2$, the opposite happens, and the indifference curve gets shallower.




The monotonicity axiom, therefore, can be expressed as a condition on how the slopes of the indifference curves change with respect to changes in $\mathbf{f}$. The slope of the indifference curve is the *marginal rate of attribute substitution (MRAS)*; define the function $MRAS_{\mathbf{x}}(f_1, f_2)$ as the slope of the $\mathbf{f}$ indifference curve at point $\mathbf{x}$[14]. *Compromise/Attraction Monotonicity* is therefore equivalent to:

$$\frac{\partial}{\partial f_1} MRAS_{\mathbf{x}}(f_1, f_2) > 0$$
$$\frac{\partial}{\partial f_2} MRAS_{\mathbf{x}}(f_1, f_2) < 0$$

The $MRAS$ is equal to the ratio of the marginal utilities, so $MRAS_{\mathbf{x}}(f_1, f_2) = \frac{U_1(x_1,x_2,f_1,f_2)}{U_2(x_1,x_2,f_1,f_2)}$. Taking these derivatives shows that $U_{13}U_2 - U_1U_{23} > 0 > U_{14}U_2 - U_1U_{24}$, which is equivalent to Property 2[15]:

$$\frac{U_{13}}{U_{23}} > \frac{U_1}{U_2} > \frac{U_{14}}{U_{24}}$$

Transitivity of the pairwise preference is equivalent to Property 3:

$$\left.\begin{array}{r} U(x_1,x_2,x_1,y_2) = U(y_1,y_2,x_1,y_2) \\ U(y_1,y_2,y_1,z_2) = U(z_1,z_2,y_1,z_2) \end{array}\right\} \Rightarrow U(x_1,x_2,x_1,z_2) = U(z_1,z_2,x_1,z_2)$$

Finally, *Substitutability* is equivalent to Property 4. Given a frame $\mathbf{f}$, for each $\mathbf{x} \in A^{\mathbf{f}}$ there exists a $\mathbf{y} \in A^{\mathbf{f}}$ such that

$$U(x_1,x_2,f_1,f_2) = U(f_1,y_2,f_1,f_2) = U(y_1,f_2,f_1,f_2)$$

This then concludes the proof that a representation satisfying properties 1-4 exists *if* Axioms 1-6 are satisfied; it remains to be shown that if the representation exists, it satisfies the axioms.

Given a representation $U(\mathbf{x}, \mathbf{f})$, the usual argument implies *Continuous Weak Order*. Being strictly increasing in $x_1$ and $x_2$ implies *Simplicity*. Property 4 implies *Substitutability*. Property 3 implies Pairwise Transitivity, and Property 2 implies *Compromise/Attraction Monotonicity*. These are all trivial. The remaining axioms (*Frame Continuity* and Pairwise Continuity) require more sophisticated arguments.

Consider $\{(\mathbf{y}^n)\}_{n=1}^{\infty} \to (\mathbf{y})$, where $\mathbf{x} \sim^* (\mathbf{y}^n)$ for each $n$. Pairwise Continuity is implied if $\mathbf{x} \sim^* (y_1, y_2)$. By the continuity of $U$, $U(x_1,x_2,x_1,y_2^n) \to U(x_1,x_2,x_1,y_2)$ (Assuming, without loss of generality, that $\mathbf{x} \succ_2 \mathbf{y}$). Also by continuity, $U(y_1^n,y_2^n,x_1,y_2^n) \to U(y_1,y_2,x_1,y_2)$[16]. However, $U(x_1,x_2,x_1,y_2^n) =$

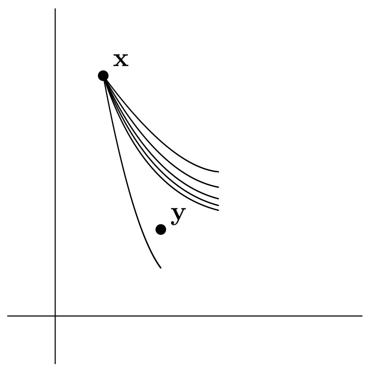
[14] This does assume that the indifference curve has a slope at point $\mathbf{x}$; however, note that by *Simplicity*, the indifference curves are monotonic. Therefore, they are differentiable almost everywhere. For points in the measure zero set which have no derivative, left and right derivatives can be taken to establish the property.

[15] Assuming $U_{23}$ and $U_{24}$ are not equal to zero; if they are, consider Property 2 to be the non-reduced form.

[16] This does require $U$ to be *jointly* continuous in all arguments; it can be seen that this is implied by the continuity in each individual argument as follows: to show $\{(x_1^n,x_2^n,f_1^n,f_2^n)\}_{n=1}^{\infty} \to (x_1,x_2,f_1,f_2) \Rightarrow \{U(x_1^n,x_2^n,f_1^n,f_2^n)\}_{n=1}^{\infty} \to U(x_1,x_2,f_1,f_2)$,

$U(y_1^n, y_2^n, x_1, y_2^n)\ \forall\ n$. This implies that $U(x_1, x_2, x_1, y_2) = U(y_1, y_2, x_1, y_2)$, as desired.

As for *Frame Continuity*, consider, $\mathbf{x}, \mathbf{y}, \mathbf{f}$ such that $\mathbf{x} \succ^{\mathbf{f}} \mathbf{y}$, and $f_i'$ such that $\mathbf{y} \succ^{(f_{-i}, f_i')} \mathbf{x}$. Suppose there does not exist an $f_i''$ such that $\mathbf{x} \sim^{(f_{-i}, f_i'')} \mathbf{y}$. In other words, while there is no $f_i$ which makes $\mathbf{x} \sim^{\mathbf{f}} \mathbf{y}$, there are $f_i$'s which make it both preferred and dispreferred. So as $f_i$ is lowered and the indifference curve rotates through $\mathbf{x}$ as per *Compromise/Attraction Monotonicity*, there is a gap in the area covered by the indifference curves, and $\mathbf{y}$ resides in this gap.




This case is straightforwardly ruled out by the continuity of $U$. If such a gap exists, there exists $\{f_i^n\}_{n=1}^{\infty} \to \hat{f}_i$ such the following two statements hold:

$$U(x_1, x_2, f_{-i}, f_i^n) > U(y_1, y_2, f_{-i}, f_i^n)\ \forall\ n \tag{4}$$
$$U(x_1, x_2, f_{-i}, \hat{f}_i) < U(y_1, y_2, f_{-i}, \hat{f}_i) \tag{5}$$

However, by the continuity of $U$,

---

break up the convergent sequence into $\{x_1^n\}_{n=1}^{\infty} \to x_1$, $\{x_2^n\}_{n=1}^{\infty} \to x_2$, $\{f_1^n\}_{n=1}^{\infty} \to f_1$, and $\{f_2^n\}_{n=1}^{\infty} \to f_2$. With these constituent parts, we can see:

$$\exists\ N_1 \text{ s.t. } \quad \forall\ n_1 > N_1, \quad |w(x_1, x_2, f_1, f_2^{n_1}) - w(x_1, x_2, f_1, f_2)| < \frac{\varepsilon}{4}$$
$$\exists\ N_2 \text{ s.t. } \quad \forall\ n_2 > N_2, \quad |w(x_1, x_2, f_1^{n_2}, f_2^{n_1}) - w(x_1, x_2, f_1, f_2^{n_1})| < \frac{\varepsilon}{4}$$
$$\exists\ N_3 \text{ s.t. } \quad \forall\ n_3 > N_3, \quad |w(x_1, x_2^{n_3}, f_1^{n_2}, f_2^{n_1}) - w(x_1, x_2, f_1^{n_2}, f_2^{n_1})| < \frac{\varepsilon}{4}$$
$$\exists\ N_4 \text{ s.t. } \quad \forall\ n_4 > N_4, \quad |w(x_1^{n_4}, x_2^{n_3}, f_1^{n_2}, f_2^{n_1}) - w(x_1, x_2^{n_3}, f_1^{n_2}, f_1^{n_1})| < \frac{\varepsilon}{4}$$

By the $\Delta$-inequality, $|w(x_1^{n_4}, x_2^{n_3}, f_1^{n_2}, f_2^{n_1}) - w(x_1, x_2, f_1, f_2)| \leq |w(x_1^{n_4}, x_2^{n_3}, f_1^{n_2}, f_2^{n_1}) - w(x_1, x_2^{n_3}, f_1^{n_2}, f_2^{n_1})| + |w(x_1, x_2^{n_3}, f_1^{n_2}, f_2^{n_1}) - w(x_1, x_2, f_1^{n_2}, f_2^{n_1})| + |w(x_1, x_2, f_1^{n_2}, f_2^{n_1}) - w(x_1, x_2, f_1, f_2^{n_1})| + |w(x_1, x_2, f_1, f_2^{n_1}) - w(x_1, x_2, f_1, f_2)|$. If we choose $n_i > \max\{N_1, N_2, N_3, N_4\}\ \forall\ i \in \{1, 2, 3, 4\}$, then the RHS of the above inequality is $< \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon$.

$$\lim_{n\to\infty} U(x_1, x_2, f_{-i}, f_i^n) = U(x_1, x_2, f_{-i}, \hat{f}_i) \tag{6}$$

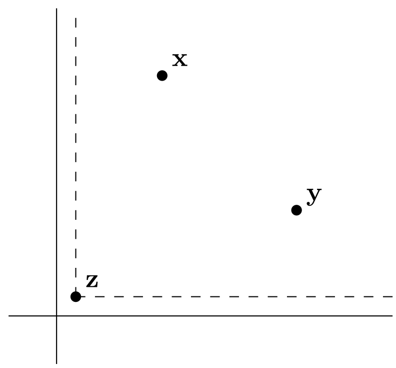
$$\lim_{n\to\infty} U(y_1, y_2, f_{-i}, f_i^n) = U(y_1, y_2, f_{-i}, \hat{f}_i) \tag{7}$$

This contradicts (4) and (5), and therefore such a gap cannot exist.

□

## 3.6 Alternate Frame Definitions

Up to this point, I have used a very specific definition of the frame. However, my representation theorem is robust to a variety of approaches to defining the frame. This is useful because there are examples where the most appropriate frame definition is not obvious, such as the following:




**z** is dominated on both attributes by both **x** and **y**. As the attraction effect is normally described in an example where the third good is only dominated by *one* other good, and thereby induces the choice of that good, it is unclear what the effect of **z** should be in this situation. The definition of the frame used up until this point suggests that it may induce a different choice, though which choice is unclear as it affects both goods. This seems plausible. One could also argue that a decision maker would completely ignore **z**, and only focus on **x** and **y**, and therefore it should have no impact. This also seems plausible. *Frame Preferences* are robust to the following definition of frame that adopts this latter interpretation of behavior:

$$\text{Define } \hat{S} \equiv \{x \in S \mid \exists\ y \in S,\ i \in \{1,2\} \text{ such that } x \succ_i y\}$$

$$\mathbf{f}(S) \equiv \left(\min_{\mathbf{x}\in\hat{S}} x_1, \min_{\mathbf{x}\in\hat{S}} x_2\right)$$

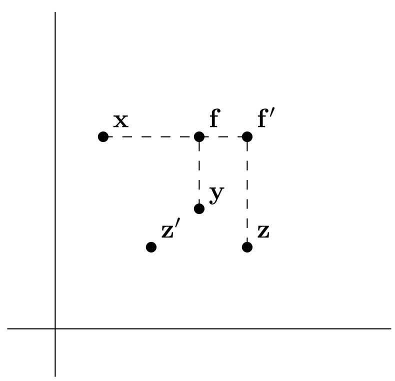
This two-step frame definition first winnows the menu by removing all points such as **z** in the picture above, which are dominated on both attributes by all other goods. It then applies the original frame definition to the winnowed menu. A collection of revealed frame preferences can be constructed using this definition of frame, and the representation theorem will still hold for this new collection of frame preferences.

In fact, any frame definition and associated collection of revealed frame preferences which satisfy the following weak condition will generate a collection of revealed frame preferences for which the representation theorem holds:

**Condition 1** (*Pairwise Edge Consistency*)**.** Given $(v_1, f_2), (f_1, v_2)$ such that $(v_1, f_2) \sim^{\mathbf{f}} (f_1, v_2)$, $\mathbf{f}\left(\{(v_1, f_2), (f_1, v_2)\}\right) = \mathbf{f}$.

This works because the proof of the representation theorem only relies on the definition of the frame in one place. After having established the representations $v_1$ and $v_2$, it is noted that, by transitivity of $\succsim^{\mathbf{f}}$, $\mathbf{a} \equiv (v_1(\mathbf{x}, \mathbf{f}), f_2) \sim^{\mathbf{f}} (f_1, v_2(\mathbf{x}, \mathbf{f})) \equiv \mathbf{b}$. Because[17] $\mathbf{f}(\{\mathbf{a}, \mathbf{b}\}) = \mathbf{f}$, this implies $\mathbf{a} \sim^* \mathbf{b}$. That is all the representation theorem requires of the frame definition; that the menu consisting solely of the two $\mathbf{f}$-indifferent goods used to identify $v_1$ and $v_2$ also have the frame $\mathbf{f}$, thus guaranteeing their pairwise indifference.

Indeed, there are many frame definitions which would generate a collection of preferences with a representation satisfying the given properties. However, many of these frame definitions would not combine with *Compromise/Attraction Monotonicity* to produce an outcome consistent with the compromise and attraction effects. For example, defining the frame as the maximum along each attribute would allow for the compromise effect, but not the attraction effect (adding $\mathbf{z}$ would shift the frame, but adding $\mathbf{z}'$ would not).




However, reference dependence abounds, and there may be other effects best illuminated through other frame definitions. Combining a frame definition and associated collection of frame preferences which satisfy *Pairwise Edge Consistency* with *Pairwise Weak Order* can be used as a basis for creating a utility function which is continuous in the frame. Specifically, if the behavior of the preferences is such that there exist representations $v_i$ along the edge of the frame as defined in the proof, and each $v_i$ is continuous in $f_i$, then *Pairwise Edge Consistency* and *Pairwise Weak Order* make those representation func-

[17]By construction, $v_1(\mathbf{x}) > f_1$, and $v_2(\mathbf{x}) > f_2$.

tions continuous in all components of $\mathbf{f}$. This is a potentially useful construct for further study of reference dependence.

## 4 Conclusion

Though the compromise and attraction effects have long been well established experimentally, there is little available in the way of systematic representations. This paper offers a fully axiomatized model which incorporates both effects, by embracing the notion of frame preferences and using them to create a succinct expression of the effects which has a natural mathematical interpretation, summarized by the *Compromise/Attraction Monotonicity* axiom and the *Compromise/Attraction Rotation* property.

Through its regularity properties, this model is very convenient for applications. It is natural to consider compromise and attraction effects in an industrial organization setting, for example in the case of a multi-product monopolist, or a single product oligopoly game. Ok et al. (2011) do the former, using their model which features the attraction effect but not the compromise effect. The latter situation remains unexplored. In both cases, this model would be ideal for exploring optimal strategies in the presence of both effects.

The compromise and attraction effects can also be applied to the political realm. Pan et al. (1995) show evidence of the attraction effect in voting results from the 1992 U.S. presidential election and the 1994 Illinois Democratic gubernatorial primary. Relatedly, it is a commonly observed phenomenon that many elections feature extremist candidates with no possibility of winning. While no other papers provide explanations for this result, Poterack and Solow (2015) use Frame Preferences to do so.

In addition, this paper makes behavioral predictions which can be tested experimentally. In particular, while papers have studied both compromise and attraction effects, there has not been an attempt to construct a single experiment which can show either effect in the same setting. Such an experiment could be used to estimate the relative magnitudes of the effects, which this paper predicts would be the same. Also, the effect of adding goods which do not change the frame should be explored; i.e., it should be tested whether the addition of non-extreme goods can induce either effect.

Finally, as noted the paper draws a relationship between *Pairwise Weak Order* and a representation which is continuous in the frame. As reference dependence is common, and as constructing a utility which is continuous in the reference is desirable, this is a potentially useful basis for future papers looking to construct convenient reference dependent representations.

## A The $n$-attribute Case

For simplicity, the main body of the paper assumes there are only two attributes along which goods are judged. However, the model extends naturally to the $n$-

attribute case, with minimal modification. Suppose the set of goods is $\mathbb{R}^n$. As before, define attribute preferences $\succsim_1, \ldots, \succsim_n$. The natural extension of the frame definition is $\mathbf{f}(S) = (\min_{\mathbf{x}\in S} x_1, \ldots, \min_{\mathbf{x}\in S} x_n)$. Having defined the frame, it is now easy to define the frame preference by

$$\mathbf{x} \succsim^{\mathbf{f}} \mathbf{y} \Leftrightarrow \exists\ S \in \mathcal{S} \text{ such that } \mathbf{x}, \mathbf{y} \in S,\ \mathbf{f}(S) = \mathbf{f},\ \mathbf{x} \in C(S)$$

As before, $\succsim^{\mathbf{f}}$ is only defined on

$$A^{\mathbf{f}} \equiv \{\mathbf{x} \in X | x_i \geq f_i\ \forall\ i \in \{1, \ldots, n\}\} \tag{8}$$

Finally, the definition of the pairwise preference $\succsim^*$ is unchanged: $\mathbf{x} \succsim^* \mathbf{y} \Leftrightarrow \mathbf{x} \in C(\{\mathbf{x}, \mathbf{y}\})$.

The properties have trivial extensions to the $n$-attribute case, with the exception of Property 2, which poses more of a challenge. Recall that in the $\mathbb{R}^2$ case, Property 2 is a statement about changes in the slope of the indifference curve; therefore, in the $\mathbb{R}^n$ case, it is a statement about changes in the norm of the hyperplane tangent to the indifference surfaces.

By Property 2, lowering the frame makes the indifference curves steeper. Another way to express that is when the frame is decreased in the second component, a vector perpendicular to an indifference curve at a given point will also decrease in the second component. Therefore, the $n$-dimensionally equivalent statement is that when the frame is decreased in the $i$th component, the norm of the hyperplane tangent to the indifference surface *also* decreases in the $i$th component.

Define $u^{\mathbf{f}}(\mathbf{x}) \equiv U(\mathbf{x}, \mathbf{f})$. $\nabla u^{\mathbf{f}}(\mathbf{x}) = \langle u_1^{\mathbf{f}}(\mathbf{x}), \ldots, u_n^{\mathbf{f}}(\mathbf{x})\rangle$ is the norm of the hyperplane tangent to the indifference surface on which $\mathbf{x}$ lies. The desired property is that if $\mathbf{f}' \equiv (\mathbf{f}_1, \ldots, \mathbf{f}_i - \delta, \ldots, \mathbf{f}_n)$, for some $\delta > 0$, then $\exists\ \varepsilon > 0$ such that

$$\frac{\nabla u^{\mathbf{f}'}(\mathbf{x})}{||\nabla u^{\mathbf{f}'}(\mathbf{x})||} = \frac{\langle u_1^{\mathbf{f}}(\mathbf{x}), \ldots, u_i^{\mathbf{f}}(\mathbf{x}) - \varepsilon, \ldots, u_n^{\mathbf{f}}(\mathbf{x})\rangle}{||\langle u_1^{\mathbf{f}}(\mathbf{x}), \ldots, u_i^{\mathbf{f}}(\mathbf{x}) - \varepsilon, \ldots, u_n^{\mathbf{f}}(\mathbf{x})\rangle||}$$

i.e, the norm of the new hyperplane tangent to the indifference surface on which $\mathbf{x}$ lies after the frame is lowered on the $i$th component to $\mathbf{f}'$ overlaps the old one decreased on the $i$th component by some amount.

As with the properties, the axioms remain either entirely unchanged, or have trivial extensions to the $n$-attribute case, with the exception of *Compromise/Attraction Monotonicity*. The extension of this axiom requires some care; however, upon reflection it should seem quite natural and intuitive, and clearly equivalent to the $n$-dimensional statement of Property 2.

- *Compromise/Attraction Monotonicity*: Given $\mathbf{x} \sim^{\mathbf{f}} \mathbf{y}$, define $A, B, \subset \{1, \ldots, n\}$ s.t. $i \in A \Rightarrow \mathbf{x} \succ_i \mathbf{y}$, $j \in B \Rightarrow \mathbf{y} \succ_j \mathbf{x}$.

  1. For each $i \in A$,
     - $\mathbf{y} \succ^{(f_{-i}, f_i')} \mathbf{x}\ \forall\ f_i' < f_i$
     - $\mathbf{x} \succ^{(f_{-i}, f_i'')} \mathbf{y}\ \forall\ f_i'' \in (f_i, y_i]$

2. For each $j \in B$,
   - $\mathbf{x} \succ^{(f_{-j}, f'_j)} \mathbf{y} \;\forall\; f'_j < f_j$
   - $\mathbf{y} \succ^{(f_{-j}, f''_j)} \mathbf{x} \;\forall\; f''_j \in (f_j, x_j]$
3. For each $k \in (A \cup B)^c$,
   - $\mathbf{x} \sim^{(f_{-k}, f'_k)} \mathbf{y} \;\forall\; f'_k$

Now with the axioms sorted, proceed as before, proposing a representation. Define $v_i(\mathbf{x}, \mathbf{f})$ as the solution to

$$\mathbf{x} \sim^{\mathbf{f}} (f_1, \ldots, f_{i-1}, v_i(x_1, \ldots, x_n, f_1, \ldots, f_n), f_{i+1}, \ldots, f_n)$$

$v_i$ is a representation. It is strictly increasing and continuous in $x_j$, $\forall\; j$. $v_i$ is decreasing and continuous in $f_i$. Given $f'_i > f_i$, define $\mathbf{f}' \equiv (f_1, \ldots, f'_i, \ldots, f_n)$, and

$$(f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}), \ldots, f_n) \succ^{\mathbf{f}'} (f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}'), \ldots, f_n)$$

This is because it must be the case that $v_i(\mathbf{x}, \mathbf{f}) \succ_i \mathbf{x}$. Now, consider $f_i^m \to f_i$. Without loss of generality, suppose this is increasing. Define $\mathbf{f}^m \equiv (f_1, \ldots, f_i^m, \ldots, f_n)$. $\mathbf{f}^m \to \mathbf{f}$. Because $v_i$ is decreasing in $f_i$,

$$v_i(\mathbf{x}, \mathbf{f}) \leq v_i(\mathbf{x}, \mathbf{f}^m) \;\forall\; m$$

Because $v_i(\mathbf{x}, \mathbf{f}^m)$ is bounded from below, $\exists\; \inf_m\{v_i(\mathbf{x}, \mathbf{f}^m)\}$, and $v_i(\mathbf{x}, \mathbf{f}) \leq \inf_m\{v_i(\mathbf{x}, \mathbf{f}^m)\}$. If this inequality were strict, it would violate *Frame Continuity*, so it must hold with equality, which implies continuity in $f_i$.

Now consider $\{(f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}), \ldots, f_n)\}_{i=1}^n$. All of these points are $\mathbf{f}$-indifferent to $\mathbf{x}$, so they are all $\mathbf{f}$-indifferent to each other. Furthermore, for any two of them, the frame of just the pair is $\mathbf{f}$. So any two of the points are pairwise indifferent. By *Pairwise Weak Order*, $\exists\; u^*$ representing the pairwise preference. $u^*$ can show $v_i$ is decreasing and continuous in $f_j$.

Consider $\{f_j^p\}_{p=1}^\infty \to f_j$. ($\mathbf{f}^p = (f_1, \ldots, f_j^p, \ldots, f_n)$). Because $v_j$ is continuous in $f_j$, $\{v_j(\mathbf{x}, \mathbf{f}^p)\}_{p=1}^\infty \to v_j(\mathbf{x}, \mathbf{f})$. $\forall\; p$,

$$\begin{aligned} &u^*(f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}^p), \ldots, f_j^p, \ldots, f_n) \\ =&u^*(f_1, \ldots, v_j(\mathbf{x}, \mathbf{f}^p), \ldots, f_n) \end{aligned}$$

and

$$\begin{aligned} &u^*(f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}), \ldots, f_n) \\ =&u^*(f_1, \ldots, v_j(\mathbf{x}, \mathbf{f}), \ldots, f_n) \end{aligned}$$

By continuity of $u^*$,

$$\{u^*(f_1, \ldots, v_j(\mathbf{x}, \mathbf{f}^p), \ldots, f_n)\}_{p=1}^\infty \to u^*(f_1, \ldots, v_j(\mathbf{x}, \mathbf{f}), \ldots, f_p)$$

and therefore

$$\{u^*(f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}^p), \ldots, f_j^p, \ldots, f_n)\}_{p=1}^\infty \to u^*(f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}), \ldots, f_n)$$

Therefore, $U = u^*(f_1, \ldots, v_i(\mathbf{x}, \mathbf{f}), \ldots, f_n)\ \forall\ i$ satisfies Property 1. Property 2 is implied by the $n$-dimensional version of *Compromise/Attraction Monotonicity*. Properties 3 and 4 are still implied by Pairwise Transitivity and *Substitutability*, respectively.

To show the representation implies the axioms, it remains that the usual argument implies *Continuous Weak Order*, Property 4 implies *Substitutability*, Property 3 implies Pairwise Transitivity, Property 2 implies *Compromise/Attraction Monotonicity*, and strict increasing in the first $n$ arguments implies *Simplicity*. This leaves *Frame Continuity* and *Pairwise Continuity*, and the two-dimensional proofs of these generalize easily to $n$ dimensions.

# B Non-monotonic Attribute Preferences

In many applications, it is natural to have a good's desirability increase monotonically in an attribute. For example, all else equal, computers with more RAM, cars with more gas mileage, and televisions with better picture quality are all preferable. However, when considering the number of ports on a computer, or the color of a car, or the size of a television, it is not obvious that preferences over these attributes can be mapped to a monotonically increasing component.

However, attribute preferences can be established with the following properties, which are implied by the condition used in the paper:

1. Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, if there exists $p \in \mathbb{R}^{n-1}$ such that $(x_i, p) \in C\left(\{(x_i, p), (y_i, p)\}\right)$, then for each $p' \in \mathbb{R}^{n-1}$, $(x_i, p') \in C\left(\{(x_i, p'), (y_i, p')\}\right)$.

2. "Attribute WARP (aWARP):" Given $S, S' \in \mathcal{S}$, $i \in \{1, \ldots, n\}$ such that $\forall\ s \in S, S'$, $s_{-i} = p_{-i}$ $(\forall\ -i \neq i)$; if $\mathbf{x}, \mathbf{y} \in S \cap S'$ and $\mathbf{x} \in C(S)$, then $\mathbf{y} \in C(S')$ implies $\mathbf{x} \in C(S')$.

3. $\{y \in \mathbb{R} | y \in C\left(\{(x, p), (y, p)\}\right)\}$ and $\{z \in \mathbb{R} | x \in C\left(\{(x, p), (z, p)\}\right)\}$ are both closed.

The first property establishes that the attributes are indeed distinct; if $x_i \succ y_i$ when they share some common set of other attribute values $p$, it remains so for any other set of attribute values $p'$. The second is a variant on WARP, applying only to cases where the differences exist only along one attribute. The final property ensures continuity.

Now, define $\mathbf{x} \succsim_i \mathbf{y}$ if and only if $(x_i, p) \in C\left(\{(x_i, p), (y_i, p)\}\right)$ for some $p \in \mathbb{R}^{n-1}$. $\succsim_i$ is trivially complete. aWARP implies it is also transitive. Suppose $\mathbf{x} \succ_i \mathbf{y}$, $\mathbf{y} \succ_i \mathbf{z}$.

$$
\begin{aligned}
&(x_i, p) \in C\left(\{(x_i, p), (y_i, p)\}\right) \\
&(y_i, p) \in C\left(\{(y_i, p), (z_i, p)\}\right) \\
\Rightarrow\ &(y_i, p) \in C\left(\{(x_i, p), (y_i, p), (z_i, p)\}\right) \\
\Rightarrow\ &(x_i, p) \in C\left(\{(x_i, p), (y_i, p), (z_i, p)\}\right) \\
\Rightarrow\ &(x_i, p) \in C\left(\{(x_i, p), (z_i, p)\}\right) \\
&\Rightarrow \mathbf{x} \succ_i \mathbf{z}
\end{aligned}
$$

Finally, because property 3 ensures $\succ_i$ is continuous, $\exists\ u_i$ representing $\succ_i$ $\forall\ i$. We can now define a mapping $M : \mathbb{R}^n \to \mathbb{R}^n$ by $M(x) = (u_1(x), \ldots, u_n(x))$. This maps the goods to a space where their desirability is indeed increasing in the attributes, and the rest of the proof may proceed as written. These utility functions are not necessarily surjective, so the space of goods may not cover all of $\mathbb{R}^n$, but that is not a problem. This technique assumes that two goods $\mathbf{x}$ and $\mathbf{y}$ such that $\mathbf{x} \sim_i \mathbf{y}\ \forall\ i$ are treated identically.

# References


Barbos, A. (2010). Context effects: A representation of choices from categories. *Journal of Economic Theory*, 145:1224–1243.

de Clippel, G. and Eliaz, K. (2012). Reason-based choice: A bargaining rationale for the attraction and compromise effects. *Theoretical Economics*, 7:125–162.

Huber, J., Payne, J. W., and Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 9:90–98.

Kivetz, R., Netzer, O., and Srinivasan, V. (2004). Alternative models for capturing the compromise effect. *Journal of Marketing Research*, 41:237–257.

Ok, E., Ortoleva, P., and Riella, G. (2011). Theory of product differentiation in the presence of the attraction effect. Mimeo, New York University.

Ok, E., Ortoleva, P., and Riella, G. (2015). Revealed (p)reference theory. *American Economic Review*, 105(1):299–321.

Pan, Y., O'Curry, S., and Pitts, R. (1995). The attraction effect and political choice in two elections. *Journal of Consumer Pyschology*, 4(1):85–101.

Poterack, A. and Solow, B. (2015). Extremist politics and the preference for compromise. Mimeo, Boston University.

Ravid, D. (2015). Focus, then compare. Mimeo, University of Chicago.

Rubinstein, A. and Salant, Y. (2008). (a,f), choice with frames. *Review of Economic Studies*, 75:1287–1296.

Simonson, I. (1989). Choice based on reasons: The case for the attraction and compromise effects. *Journal of Consumer Research*, 16:158–174.

Simonson, I. and Tversky, A. (1993). Context-dependent preferences. *Management Science*, 39:1179–1189.

Tserenjigmid, G. (2015). Choosing with the worst in mind: A reference-dependent model. Mimeo, California Institute of Technology. First draft 2014, current version 2015.

Tversky, A. and Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4):1039–1061.