LOOK BEFORE YOU LEAP: CONTINUOUSLY EVOLVING UTILITY REPRESENTATIONS AND THE ATTRACTION EFFECT

by

Christian Arthur Fagan

Submitted to the Undergraduate Faculty

of the School of Arts and Sciences in partial fulfillment of

the requirements for the degree of Bachelor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Christian Arthur Fagan

It was defended on

April 19th, 2012

and approved by

Roee Teper, PhD, Assistant Professor, University of Pittsburgh, Department of Economics

Sourav Bhattacharya, PhD, Assistant Professor, University of Pittsburgh, Department of Economics

Juan Sebastian Llerás, PhD, Postdoctoral Researcher, Carnegie Mellon University, Department of Social and Decision Sciences

Thesis Advisor: Luca Rigotti, PhD, Associate Professor, University of Pittsburgh, Department of Economics

LOOK BEFORE YOU LEAP: CONTINUOUSLY EVOLVING UTILITY REPRESENTATIONS AND THE ATTRACTION EFFECT

Christian Fagan

University of Pittsburgh, 2012

We study a decision-maker who is presented with a menu of three options to consume. The options are such that two of the bundles are extremely similar, though one is demonstrably inferior to the remaining options. We propose and analyze a model in which the decision-maker's preferences change continuously as he is initially presented with information before learning his true preferences. The decision-maker initially exhibits what is known as the "attraction effect," whereby he is influenced by the similarity in his choice set. As time passes the attraction effect diminishes and the decision-maker's decision-making is aligned with the Tversky hypothesis, the antithetical counterpart to the attraction effect. The decision-maker's optimal behavior depends on his patience and capacity for learning in addition to his baseline preferences.

Keywords: attraction effect, similarity hypothesis, preference reversal, cost of thinking, homeomorphism, homotopy

Contents

1 Introduction

Modern microeconomic theory seeks to ascribe individual decision-making to well-behaved, axiomatic representations of preferences. However, it is readily apparent that decision-making is context dependent. The manner in which people perceive things through framing effects and other situational variables has a drastic effect on their choices. It is possible to account for such behavior through relaxation of the weak axiom of revealed preference, through probabilistic methods of choice, and several other methods.

In this paper, we propose a model of intertemporal utility maximization in which a decision-maker's choices are influenced by a few key factors as time passes. The decision-maker has ex ante preferences that are influenced directly by the framing effect known as the "attraction effect." The attraction effect creates an initial distortion of the consumer's preferences, and this effect is ameliorated by the consumer's ability to learn his true preferences as time elapses. As time goes on further, the consumer begins to exhibit the effects described by Amos Tversky (1972) in his seminal work on the similarity puzzle. The Tversky hypothesis proposes the polar opposite of the empirically supported attraction effect. Ultimately, the decision-maker's optimal behavior in the context of the model will reflect the empirical observations in these settings. Our main contribution is that we provide a model of utility maximization that accounts for the decision-makers continuously evolving preferences without relaxing the weak axiom or incorporating random choice behavior.

2 Literature Review

The initial work on discrepancies between patterns in individual choice and those predicted by seminal microeconomic theory can be attributed to Luce (1959) and a subsequent response by Debreu (1960). In his paper, Luce proposes a "random choice hypothesis that claims:

> The ratio between the probability with which option j is chosen from a set of options to the probability with which k is chosen from the same set is constant across all sets that contain j and k.

Luce uses this hypothesis to provide a model of random choice. However, Debreu noted a key flaw in the design of Luce's model. Debreu identified what would come to be known as "the duplicates problem." Imagine a world in which a decision-maker is indifferent between a train and a bus that happens to be blue. In this setting, the decision-maker chooses either with equal probability of one half. Suppose also that the decision-maker has no preferences over the color of a bus, so that he also is indifferent among riding a blue bus or a yellow bus. Since the decision-maker is utterly indifferent with respect to color, we should expect that he would still elect to ride the train with probability one half should he be faced with these three options. But in Luce's model of random choice, his probability of selecting to ride the train would be reduced to one third.

Debreu's example can be considered extreme in some sense. People of course have idiosyncratic preferences for minor details like color. Even so, it does illustrate a drawback of the Luce model. It cannot account for the manner in which decision-makers identify options as substitutes if not perfect duplicates. In order to account for such behavior, famed behavioral research Amos Tversky proposes in (1972b) what he refers to as the similarity hypothesis:

> The addition of an alternative to an offered set 'hurts' alternatives that are similar to the added alternative more than those that are dissimilar to it.

There are a number of ways to interpret this effect in the context of choice behavior. It reflects the idea that a decision-maker's preferences are often independent of slight differences between options. It also has a more psychological interpretation when compared to the attraction effect. The attraction effect seems to occur because the pairing of similar options provides the decision-maker with a convenient point of comparison when one is demonstrably better than another. Alternatively, the Tversky similarity hypothesis paints a picture whereby similar options are associated together, or categorized in some sense. In this light, an option that the decision-maker originally deems fit for selection may seem weaker when he categorizes it as similar to a significantly weaker one

Empirical evidence lends heavy support for the attraction effect. Experiments originating back to Huber, Payne and Puto (1982) have all pointed to the existence of the attraction effect in settings that include voting on political candidates, financial investment and even medical decisions. It is plain that the attraction effect and the Tversky similarity hypothesis are utterly incompatible. Whereas the Tversky similarity hypothesis can be thought of as a measure to account for a decision-maker's psychological associations while maintaining the well-behaved notion of preference monotonicity, the observed attraction effect utterly shatters this idea. This is particularly problematic for random utility models that rely on the notion of monotonicity in a probabilistic setting.

The work of Gul, Natenzon and Pesendorfer (2012) and particularly Natenzon (2010) features random choice models that can account for violations of monotonicity. Whereas previous models attribute the attraction effect to a lapse in attentiveness or capability on the behalf of the decision-maker, the Bayesian probit process described by Natenzon allows for the existence of a decision-maker who is ultimately aware of his options. The decision-maker makes inferences about his true utility by updating his beliefs according to Bayes rule as time passes. In the scope of the Bayesian probit process, the capacity for learning is homogeneous across decision-makers. Instead of having different learning rates, the precision of a decision-maker's information depends only on how much time has elapsed since the process began. Information becomes arbitrarily precise as time tends to infinity. In this model, learning can be endogenized so that a decision-maker's optimal behavior depends upon the amount of time it takes him to learn his preferences.

The Bayesian probit process also relies heavily on the notion of similarity. In the context of that model, similarity rises according to the correlation parameter of the standard Gaussian distribution. In this model, similarity is simply determined by Euclidean distance. Bundles that are close together are similar and vice versa. Similarity as correlation induces the attraction effect in the early stages of the Bayesian probit process, since the decision-maker's ranking for similar options is much easier to discern. But as time passes, the attraction effect diminishes entirely, giving way to the effects described by Tversky's similarity hypothesis. In the limit, Natenzon's decision-maker can discern perfectly between trains, green buses, and yellow buses as choice probabilities converge to reflect the decision-maker's indifference among these options.

Another important recent paper that describes the attraction effect is DeClippel and Eliaz (2012). These two took an entirely different approach to explaining the attraction effect. The authors note that decision-making occurs across a number of dimensions. To hearken back to the example above, people may have preferences not only over types of transportation, but over other facts like color or safety. They establish a model based on what Tversky et al (1992, 1993) refer to as "reason-based choice." That is, choice behavior may be explained as the decision-maker weighing various attributes on the spot. DeClippel and Eliaz model this process as a cooperative dual-selves bargaining model. In the scope of their model, the Nash bargaining or "fallback" solution is the only axiomatic cooperative bargaining solution that satisfies the necessary conditions for the attraction effect.

3 Preliminary Examples

There are demonstrably a number of methods by which the attraction effect can be explained. One reason for this is that there a plethora of exceptionally clever people who work on these problems. Another is that because of the idiosyncratic nature of human behavior and decision-making, there is no true consensus on why the attraction effect occurs. This model will draw heavily from the intuition of Natenzon (2010). The fact that decision-making is often lengthy and quite costly lends credence to the idea that decision-maker should be susceptible to the attraction effect early in the process. As alluded to previously, this model will not use random utility maximization or axiomatic bargaining to account for the attraction effect. Instead, the decision-maker's utility will be represented at different times as a sequence of functions. In this setting, optimal behavior isn't solely a function of the precision of information, but it has a temporal element as well. The attraction effect is exhibited in many settings where intertemporal choice is relevant to examine. For example, consider a student in an economics laboratory experiment. While it is difficult to imagine that the student cares deeply about the choices that he makes within the confines of the experiment, it is rather easy to imagine that he has some preference for choosing "correctly," even if the marginal benefit of a correct decision is minimal. It is even easier to imagine that the student wants to make the decision in a timely fashion in order to substitute time spent in the lab for time doing just about anything else.

In the model, most of the attention will be drawn to menus that are subsets of $\mathbb{R}^2$. Traditionally, the attraction effect is observed in settings with menus with 3 elements, and we will stick to this convention. I will present a series of increasingly complex models that account for the attraction effect by changing the decision-maker's utility function in the presence of information. Ex ante, the decision-maker has a well defined preference relation with a continuous utility representation. In some sense, this is similar to the idea of a random utility model in which the decision-maker's true preferences are unknown to the decision-maker. In that setting, preferences are inferred by the decision-maker examining draws from a distribution and forming posterior beliefs about his true utility. The example below can be thought of in some sense as a naive process where the consumer undertakes one round of updating and then makes a decision.

4 Motivating Example (One)

There are an immense number of methods for describing consumer behavior with regards to the attraction effect. Notably, through noncooperative dual selves bargaining models and also by constructing random utility models. In this example, we will examine the attraction effect in a different context. Unlike the bargaining models, we focus on a single consumer. Unlike the random utility maximization models, the consumer's preferences are already well known. However, as in the other models, this consumer will suffer from the attraction effect because of the similarity between a maximal and an inferior bundle.

The consumer's rational preference relation $\succeq$ is a complete, transitive binary relation on $\mathbb{R}^2$ (which can be thought of as different goods, or different dimensions of one good, etc.) which is represented by a continuous, monotone utility function U. In order to account for the attraction effect, we will endow the consumer with additional utility based on the similarity between bundles offered in a given menu. This additional utility is derived from the aptitude with which similarity allows one to make a "correct" decision, particularly in settings where acquiring precise knowledge by investing in information is costly.

Let a menu M be a subset of $\mathbb{R}^2$ that contains 3 elements. We will examine such menus in order to isolate the attraction effect, which has been demonstrated to occur in this setting. Since M is finite, it is compact. Therefore, since U is continuous, U attains a maximum on M. Let d denote the Euclidean metric on $\mathbb{R}^2$, i.e.

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

Since $M \subset \mathbb{R}$, M inherits the metric d. Additionally, d is continuous on $\mathbb{R}^2$, and thus attains a minimum on M. It is this minimal distance that we will examine below.

Example: Consider $\succeq$, U and d as above for a menu $M := \{A, B, C\}$ such that

$$A \sim B \succ C$$

and therefore

$$U(A) = U(B) > U(C)$$

So that $C_{\succeq}(M) = A, B$. Suppose further that

$$\min\{d(A, C), d(A, B)\} \geq d(B, C).$$

That is, B and C are closer together than any other pair of bundles in M. In the classical sense of utility maximization, it is clear that the consumer can do so by randomizing between selecting A or B using any probability distribution. In this model, when a consumer observes a menu, his utility is mapped to a new utility function $H : U(M) \to R$ where

$$H(U(x)) = \begin{cases} U(x) + \frac{1}{d(x,C)} & \text{if } x \in \{A, B\} \\ U(x) \text{ if } x = c \end{cases}$$

It is important to note that by convention, $d(x, x) = 0$ in any metric space. We will restrict our attention to elements that were considered maximal before the consumer's utility function changes due to the presence of a bundle that facilitates comparison, so $H(C) = U(C)$ to preserve that ranking. This aligns with the spirit of other attraction effect models. The addition of $\frac{1}{d(x,y)}$ indicates the additional utility a consumer receives from having a point of comparison. If we are to think of bundles that are"close" with respect to the metric as having similar traits, then the closer an inferior bundle is to a maximal bundle, the more accurate the comparison should be. The more accurate the comparison, the more the consumer stands to gain. Since C is inferior to both A and B, it should remain inferior. Traditionally, the inclusion of C induces the consumer to select B from his choice set. The inferior bundle C remains inferior.

Now, define $\bar{U} : M \to R$ as

$$\bar{U} := H(U(x))$$

As defined above, H is the sum of continuous functions, and is thus continuous. U is continuous by assumption. Hence $\bar{U}$ is continuous, represents a continuous rational preference relation $\bar{\succsim}$ and attains a maximum on M. Since

$$d(A, C) \geq d(B, C) \iff \frac{1}{d(A, C)} \leq \frac{1}{d(B, C)}$$

and

$$U(A) = U(B)$$

it is evident that

$$\bar{U}(B) \geq \bar{U}(A)$$

and B is the unique element in $C_{\bar{\succsim}}(M)$.

5 Motivating Example (Two)

The above example is, of course, primitive. It does account for the attraction effect; once the list of alternatives is presented to the consumer, his choice behavior is affected by similarity. C is closer to B than to A and thus more similar, and this is the driving force behind the change in utility. Since this model is simplistic, there are few shortcomings. One is that, ideally, there would be a way to map $U(C)$ to another value. However, making $U(C)$ a fixed point is somewhat intuitive. There is no evidence from lab or field data to support the notion that the utility the consumer derives from the initially dominated option should change in any way. We can trust our consumer to recognize its inferiority to the remaining options. There are other shortcomings of this simple model. The most obvious one is that it ignores how the consumer's preferences will change over time. Since this is an analog to one round of updating, we require an analogy to a random utility on an ordered time set. We will accomplish this again by a device that maps the consumer's existing utility into a new set. One thing that random utility does not address is the cost of decision-making. A random utility-maximizing agent simply updates his beliefs about his preferences, and his information becomes more precise as time tends to infinity. In this version of the model, the agent is endowed with a discount factor $\delta \in (0, 1)$. As time passes, the consumer discounts consumption. However, the consumer stands to gain from the passage of time in a different sense. As time passes, the consumer can better learn about the alternatives that he faces. For the sake of simplicity, let the ordered time set be N. We model the consumer's preference for clarity and more precise information as an increasing, continuously differentiable function $f : N \to \mathbb{R}_+$. Adding to the model above, a general picture of what the consumer's preferences might look like in this setting is:

$$\Omega(t, x) = \delta^t f(t)H(x) = \delta^t f(t)(U(x) + \frac{1}{d(x, C)})$$

It may seem at first that this functional form accounts for the consumer's tradeoff between time and information, it isn't entirely faithful to the spirit of the attraction effect. It is easy to see that on any interval on which $\delta^t f(t)$ is increasing, $H(x)$ will scale Ω so that there is a wider gap between $\Omega(t, A)$ and $\Omega(t, B)$. This effectively *exacerbates* the attraction effect. This can be thought of as an analog to initial belief formation in which the consumer is influenced by the focal bundle C. In order to maximize Ω, we can simply examine the time dimension since H is maximized on a finite set (even though Ω is multiplicatively seperable, there is no need to examine Hessians, etc. to determine the nature of critical points). Let

$$y = \arg \max_{x \in \{A,B,C\}} H(x)$$

Then

$$\Omega_t(t, y) = (\delta^t f(t))' H(y) = (\delta^t \ln(\delta) f(t) + \delta^t f'(t))H(y) = \delta^t H(y)(f'(t) + \ln(\delta) f(t))$$

Since f is everywhere positive strictly increasing, f' is positive. So Ω is increasing when

$$\frac{f'(t)}{f(t)} > -\ln(\delta)$$

and decreasing when

$$\frac{f'(t)}{f(t)} < -\ln(\delta)$$

Given the vagueness of the functional form, these inequalities seem a bit trivial. But there are a wealth of functional forms for f that will induce Ω to be either a) strictly decreasing, or b) increasing on $[0, t_0]$ and decreasing on $[t_0, \infty)$ for some critical t_0. This is true of any polynomial of finite degree, any exponential a^t such that $\delta a \leq 1$, and a host of other deterministic functions of t.

6 The Model

There are obvious drawbacks to the models presented above. While they are comprised of well-behaved, differentiable functions, one key element to the attraction effect is missing. An effective model should account not only for the initial reversal of preferences, but also for how learning can smooth away the initial reversal. In the limit, the models above will experience this to some degree. Observe that when f is continuously differentiable and increasing on R_+

$$\lim_{t\to\infty} \frac{\delta^t f(t)}{d(x, C)} = 0.$$

But U experiences the same effect. Since this is the case, whichever bundle from $\{A, B\}$ maximizes $H(x)$ will serve to maximize Ω $\forall t$. This is in direct contrast with the Bayesian probit model presented in Natenzon (18, 2010), wherein he states:

Theorem 4 (Similarity Puzzle). If $1 \sim 2 \sim 3$, then there are times T_1, $T_2 > 0$ such that introducing 3 hurts 2 more than it hurts 1 for all times before T_1 , and the reverse occurs for all times greater than T_2.

According to the Bayesian probit model, for early times the decision maker is subject to the attraction effect. But as time goes on, the decision maker's learning process is more aligned with the Tversky hypothesis. The model I detailed above can do the former but not the latter. This version will incorporate both the initial presence of the attraction effect as well as the possibility of an ensuing preference reversal.

As above, the decision maker is endowed with a rational preference relation $\succeq$ with continuous utility representation U over all possible options.

Let U^{-1} denote the level set of U, i.e. U^{-1} is the correspondence that maps a utility level to the set of alternatives that provide the decision maker with that level of utility. U_B^{-1} will denote the level set of B. The initial motivating example will serve as a baseline for the following details and the corresponding theorem. In this setting, the information provided to the decision-maker by the inferior bundle will increase the utility of options near that bundle through a continuous process. This continuous process will "stretch" the indifference curve of the option the decision maker is currently facing, mapping it farther away from the origin so that its new utility is greater than or equal to its previous utility. I will stick to the previous convention so that $A \succ C$ and $B \succ C$ with $d(B, C)$ the minimum distance between bundles. The ranking of A and B will be left ambiguous. We define $\bar{\beta}(x, \epsilon)$ to be a closed ball centered at x with radius *epsilon*. The continuous process is fleshed out by the following theorem:

Theorem 1: $h : U_B^{-1} \cap \bar{\beta}(B, d(B, C)) \rightarrow Im(h) \subseteq \partial\bar{\beta}(B, d(B, C))$ is a homeomorphism where

$$h(x) = x\frac{d(B + (d(B, C), \theta(x)), 0)}{d(x, 0)}$$

Some explanation is in order. $(d(B, C), \theta(x))$ gives the polar coordinates of a vector whose tail begins at B. The tip is at the point on the boundary of $\bar{\beta}(B, d(B, C))$ that lies on the ray that begins at 0 and passes through the point x. That ray forms an angle φ with the horizontal axis, so φ (and by extension x) will uniquely determine the angle $\theta(x)$ that this vector forms with the horizontal.

The fact that h maps to $Im(h)$ is essentially for convenience. There really isn't a convenient manner in which to state the set that h maps to, as it will be different in each case. The image of h will consist of points along the boundary of $\bar{\beta}(B, d(B, C))$. There is no precise way to state what the set looks like because of how drastically different it may look for different combinations of utility functions and $d(B, C)$ (it will vary in size, number of connected components, etc.). Ultimately, h takes points along the indifference curve given by $U(B)$, and continuously deforms the curve by projecting the points onto $\partial\bar{\beta}$.

It is important to understand that it isn't the indifference curve that is changing per se. Indeed, if we are to simply change the indifference curve by extending it farther from the origin, this makes every bundle contained in $U_B^{-1} \cap \bar{\beta}(B, d(B, C))$ worse in some sense. The distortion could make these options potentially unaffordable given the decision maker's budget constraint, and also this would create the crossing of indifference curves. Instead, the information provided by having closely related bundles induces the decision-maker to assign greater utility to these alternatives. The decision-maker now behaves as though any such bundle x lies along the indifference curve that coincides with $h(x)$.

But still, this only accounts for the initial attraction effect. The decision-maker's utility should smoothly return to the baseline level as time passes. In the scope of random utility maximization, this is analogous to the decision-maker's beliefs converging to his true utility in the limit. However, it doesn't require a grand leap of faith to imagine a situation in which the decision-maker can learn his true utility by some finite time t_0. The following theorem will illustrate this concept.

Theorem 2: Take h as above, and let I denote the identity map. Then $\eta : U_B^{-1} \cap \bar{\beta}(B, d(B,C))X[0, t_0] \to \mathbb{R}^2$ is a homotopy where

$$\eta(x,t) = \frac{t_0 - t}{t_0} h(x) + \frac{t}{t_0} I(x) = \eta(x,t) = \frac{t_0 - t}{t_0} h(x) + \frac{t}{t_0} x$$

t_0 represents the amount of time it takes the decision-maker to learn his true preferences. The process begins at $t = 0$, when the decision-maker views M and experiences the initial attraction effect. As $t \to t_0$, he assigns utility levels to the bundles in $U_B^{-1} \cap \bar{\beta}(B, d(B,C))$ that more accurately reflect his baseline preferences. At $t = t_0$, the decision-maker's bias towards B has been eliminated, and utility levels are restored to reflect U.

This is a more elegant way of describing what is at stake in Part One. We have obtained a well-behaved process by which the decision-maker's utility levels change near the "attracted" bundle B. However, in order to shift to a model in which the decision-maker continuously discounts his utility as time passes, there is one remaining facet we must describe. As it stands, the homotopy η can create a reversal of preferences. However, in a model with discounting it is plain to see that the only rational decision is to make a selection at $t = 0$ of either A or B. After $t = 0$, the utility assigned to A will decrease according to the discount rate δ. The utility assigned to B will decrease according to both δ and η.

This is where Tversky's similarity hypothesis enters the picture. Recall that in (1972), Tversky posits that "the addition of an alternative to an offered set 'hurts' alternatives that are similar to the added alternative more than those that are dissimilar to it." There is heuristic evidence that this can occur, and Natenzon addresses this notion in the Bayesian probit model. The logical sequence should look something like

1. The decision-maker views the menu M

2. The decision-maker experiences the attraction effect

3. The decision-maker learns, and the attraction effect is subdued

4. The decision-maker begins to favor A.

The reason that item 4 occurs is unclear. A popular belief is that as time passes, the decision-maker continues to associate the similar bundles B and C. However, instead of looking better by comparison, the decision-maker's association makes B seem worse than it did before because of some kind of mental compartmentalization. If we are to make sense of this notion in the scope of this model, we can do so by exhibiting a continuous deformation of U_A^{-1}. This should occur after the attraction effect has worn off. One can easily imagine a setting in which the decision-maker's utility is restored to normal levels after the attraction effect for B has worn off, whereupon the decision-maker begins to assign higher utility to A as it is starkly contrasted with the associated bundles B and C. How much the utility of A improves is closely related to the similarity (read: distance between) B and C. If C is substantially worse than B but still similar enough to induce the attraction effect, A should benefit more substantially when the decision-maker begins to associate B and C together. Hence, the formulae for homeomorphism and homotopy for U_A^{-1} will depend on the distance between B and C.

Theorem 3: 1) $h_A : U_A^{-1} \cap \bar{\beta}(A, d(B,C)) \to Im(h_A) \subseteq \partial\bar{\beta}(A, d(B,C))$ is a homeomorphism where

$$h_A(x) = x\frac{d(A + (d(B,C), \theta(x)), 0)}{d(x, 0)}$$

2) $\eta_A : U_A^{-1} \cap \bar{\beta}(A, d(B,C))X[t_1, t_2] \to \mathbb{R}^2$ is a homotopy where

$$\eta_A(x,t) = \frac{t - t_1}{t_2 - t_1}h_A(x) + \frac{t_2 - t}{t_2 - t_1}x$$

The proof of this theorem is entirely symmetric to the two preceding theorems. It is logicial that the amount by which A improves is a function of the similarity between B and C, but it also ensures that B will remain fixed according to the second transformation. Since $d(B,C)$ is the minimum distance between any of the three bundles, A and B are at least $d(B,C)$ apart. If it is the unique minimum $d(A,B)$ is of course greater, so $A \not\subseteq U_A^{-1} \cap \bar{\beta}(B, d(B,C)$ and vice versa. If $d(A,B) = d(B,C)$, then A lies on $\partial\bar{\beta}(B, d(B,C))$, and remains fixed under h. The same is true of B and h_A.

I will now adopt some minor changes in notation. Denote h as h_B, η as η_B on $[0, t_1]$ instead of t_0. The superscript t for each η denotes the value at a given time. Additionally, we will define a family of functions

$$\Omega(x,t) = \begin{cases} U(x) & \text{if } x \in \{A, C\} \text{ and } t \in [0, t_1] \\ U(\eta_B^t(B)) & \text{if } t \in [0, t_1] \\ U(\eta_A^t(A)) & \text{if } t \in [t_1, t_2] \\ U(h_A(A)) & \text{if } t \in [t_2, \infty) \\ U(x) & \text{if } x \in \{B, C\} \text{ and } t \in [t_1, t_2] \end{cases}$$

where t_1 and t_2 are given according to the preceding theorems. At $t = 0$, the attraction effect occurs, and subsequently diminishes entirely by t_1. Once the attraction effect subsides, the decision-maker perceives A to look better than it previously did, and this effect occurs until some threshold t_2. Given these effects, we will now examine the decision-maker's optimal behavior. The decision-maker will discount exponentially according to $\delta \in [0, 1]$. In this setting the exponential discount rate can be thought of as "the cost of thinking," with lower levels of δ corresponding to smaller costs and vice versa.

Theorem 4: The solution to the maximization problem

$$\max_{x \in \{A,B.C\}, t \in [0,\infty)} e^{-\delta t}\Omega(x,t)$$

is one of

1. $(B, 0)$
2. $(A, 0)$
3. (A, τ) for some $\tau \in [t_1, t_2]$

This theorem accounts for both the initial attraction effect and the ensuing Tversky effect, and these results align with empirics. Impatient people who are especially sensitive to the effect will select B from the outset. If A is drastically better than B, then not even the attraction effect can sway someone's decision. Additonally, the more patient and/or faster learning decision-makers can hold out to consume A when it becomes an even more viable option. These are precisely the decisions that we witness in empirical examples and laboratory settings.

Example 1: Consider the utility function $U(x) = \ln(x_1) + \ln(x_2)$. Take

$A = (e, 0)$, $B = (0, e)$ and $C = (0, \frac{e}{2})$. It is elementary to check that $d(B, C) = \frac{e}{2}$ is the minimum distance between bundles on M. Hence,

$$U(x) = \begin{cases} 1 & \text{if } x \in \{A, B\} \\ 1 - \ln(2) & \text{if } x = C \end{cases}$$

We will set $t_1 = 1$ and $t_2 = 2$. At $t = 0$:

$$e^0\Omega(B, 0) = U(h_B(B)) = U(B\frac{(B + d(B, C))}{d(B, 0}) = U((0, e + \frac{e}{2})) = \ln(\frac{3e}{2})$$

$$e^0\Omega(A, 0) = U(A) = 1$$

At $t_1 = 1$:

$$e^{-\delta}\Omega(B, 1) = e^{-\delta}U(B) = e^{-\delta}$$

$$e^{-\delta}\Omega(A, 1) = e^{-\delta}U(A) = e^{-\delta}$$

At $t_2 = 2$:

$$e^{-2\delta}\Omega(B, 2) = e^{-\delta}U(B) = e^{-2\delta}$$

$$e^{-2\delta}\Omega(A, 1) = e^{-\delta}U(h_A(A)) = e^{-2\delta}\ln(\frac{3e}{2})$$

It is easy to verify that the solution to the maximization problem is $(B, 0)$. This is akin to "Motivating Example One" in some sense. The options are considered equal at first, and so h_A and h_B will have identical effects. Since utility is discounted, the attraction effect occurs.

Example 2: Once again, consider the utility function $U(x) = \ln(x_1)+\ln(x_2)$. Take $A = (\frac{4e}{3}, 0)$, $B = (0, e)$ and $C = (0, \frac{e}{2})$ so that $d(B, C)$ is the minimum distance. Again, take $t_1 = 1$ and $t_2 = 2$. At $t = 0$:

$$e^0\Omega(B, 0) = U(h_B(B)) = U(B\frac{(B + d(B, C))}{d(B, 0}) = U((0, e + \frac{e}{2})) = \ln(\frac{3e}{2})$$

$$e^0\Omega(A, 0) = U(A) = \ln(\frac{4e}{3})$$

At $t_1 = 1$:

$$e^{-\delta}\Omega(B, 1) = e^{-\delta}U(B) = e^{-\delta}$$

$$e^{-\delta}\Omega(A, 1) = e^{-\delta}U(A) = e^{-\delta}\ln(\frac{4e}{3})$$

At $t_2 = 2$:

$$e^{-2\delta}\Omega(B, 2) = e^{-\delta}U(B) = e^{-2\delta}$$

$$e^{-2\delta}\Omega(A,1) = e^{-\delta}U(h_A(A)) = e^{-2\delta}\ln(\frac{4e}{3} + \frac{e}{2}) = e^{-2\delta}\ln(\frac{11e}{6})$$

In general, we can examine the behavior of $e^{-\delta t}\Omega(A,t)$ on $[t_1, t_2]$ by use of the intermediate value and max-min theorems. For the moment, we will only make use of the latter. We can guarantee that the decision-maker's utility maximizing pair will occur on $[t_1, t_2]$ as long as

$$e^{-2\delta}\ln(\frac{11e}{6}) \geq \ln(\frac{3e}{2}) \iff e^{-2\delta} \geq \frac{\ln(\frac{3e}{2})}{\ln(\frac{11e}{6})} \approx .67$$

$$\iff -2\delta \gtrapprox -.4 \iff \delta \lessapprox .2$$

For such values of δ, waiting until $t_2 = 2$ to consume is at least as good as picking B from the outset. Hence, the maximum on $[t_1, t_2]$ is the optimal selection. Note also that it is possible that larger values of δ will feature a solution of the form (A, τ) for $\tau \in [t_1, t_2]$; establishing such values would require an exhaustive solution.

This example also displays the preference reversals that this model can account for across time periods. According to the baseline preferences, A is the superior option. But when faced with a menu M, the decision-maker is influenced by a framing effect and is subjected to the attraction effect. As the decision-maker learns, his preferences return to the baseline level. But by $t_1 = 1$, utility has been discounted as the decision-maker has devoted a nontrivial amount of time to discerning his tastes. After t_1, the decision-maker's preferences begin to evolve again as A becomes favored relative to B and C even more than it was initially. At some time in $[t_1, t_2]$ the decision-maker will pick an optimal stopping time and select A whenever his discount factor δ satisfies the given inequality, though there are presumably greater discount factors for which this is also the case.

7 Conclusion

We have presented an attraction effect model that incorporates many facets of the existing literature. Whereas other models rely on concepts like axiomatic bargaining or random utility maximization, this model provides a process by which a decision-maker's preferences change continuously as time elapses. A decision-maker who is equipped with a rational preference relation and continuous utility representation is presented with a choice set that features 3 elements, including a strictly dominated option. Initially, the decision-maker assigns higher utility to the "attracted" bundle until he has a sufficient amount of time to learn his true preferences. At that time, the decision-maker begins to exhibit the effect described in the Tversky hypothesis. A logical explanation is that the decision-maker begins to associate the similar bundles in a negative light, which makes the remaining option look better by comparison. The decision-maker faces a dynamic optimization problem subject to a tradeoff between how his preferences change continuously in time and how he discounts future consumption. The solutions to the optimization problem coincide with an outcome that reflects either the attraction effect or the Tversky hypothesis, and the outcome is determined by the decision-maker's capacity for patience and effective learning.

8 Appendix

Proof of Theorem 1: First, note that U_x^{-1} is closed $\forall x$. Take $x \in S = \{y \in \mathbb{R}^2 | U(y) = \alpha\}$. Since the kernel of any continuous function is closed, $Ker(U(x) - \alpha)$ is closed. But this just yields the set S, so S is closed. Additionally, $\bar{\beta}(B, \epsilon)$ is closed by definition. Hence, $U_B^{-1} \cap \bar{\beta}(B, \epsilon)$ is an intersection of closed sets, and is thus closed. Also, this intersection is contained in $\bar{\beta}(B, \epsilon)$. Hence $U_B^{-1} \cap \bar{\beta}(B, \epsilon)$ is a closed, bounded subset of $\mathbb{R}^2$, and is thus compact. Consider the function

$$h(x) = x \frac{d(B + (d(B, C), \theta(x)), 0)}{d(x, 0)}$$

This function is the product of the identity function and the ratio of two distance functions, all of which are continuous. Hence, h is continuous. We will now show that h is injective

$$x = y \iff \frac{x}{d(x, 0)} = \frac{y}{d(y, 0)}$$

$$\iff x \frac{d(B + (d(B, C), \theta(x)), 0)}{d(x, 0)} = y \frac{d(B + (d(B, C), \theta(y)), 0)}{d(y, 0)} \iff h(x) = h(y)$$

Since the space that h maps to is $Im(h)$, it is surjective by definition. So h is both injective and surjective, and therefore bijective. Finally, note that $Im(H)$ is a subspace of $\mathbb{R}^2$, which features the standard topology according to d. Since $\mathbb{R}^2$ is Hausdorff, any subspace of $\mathbb{R}^2$ with the subspace topology is also Hausdorff. Since h is a continuous bijection from a compact set to $Im(h)$, a Hausdorff space, h is a homeomorphism.

Proof of Theorem 2: As shown in the proof of theorem 1, h is a continuous function on $\mathbb{R}^2$. Additionally, the identity function is continuous on $\mathbb{R}^2$. The functions $\frac{t}{t_0}$ and $\frac{t_0-t}{t_0}$ are ratios of linear functions, and are thus continuous. Then the restriction map $\eta_\tau : [0, t_0] \to \mathbb{R}^2$ given by

$$\frac{t_0 - \tau}{t_0} h(x) + \frac{\tau}{t_0} x$$

is continuous $\forall \tau \in [0, t_0]$ and any x on which h is defined. Hence, η satisfies the continuity conditions for homotopy.

Proof of Theorem 3: 1) As shown above, $U_A^{-1} \cap \bar{\beta}(A, \epsilon)$ is a compact subset of $\mathbb{R}^2$.. Additionally,

$$h_A(x) = x \frac{d(A + (d(A, C), \theta(x)), 0)}{d(x, 0)}$$

is a continuous bijection as shown in theorem 1. Since h_A is a continuous bijection from $\mathbb{R}^2$ to $Im(H)$, a Hausdorff subspace of $\mathbb{R}^2$, h_A is a homeomorphism.
2) Since h_A and the identity function are both continuous on $\mathbb{R}^2$, and the functions $\frac{t-t_1}{t_2-t_1}$ and $\frac{t_2-t}{t_2-t_1}$ are continuous on $[t_1, t_2]$, we observe that the restriction map $\eta_A^\tau : [t_1, t_2] \to \mathbb{R}^2$ given by

$$\frac{\tau - t_1}{t_2 - t_1} h(x) + \frac{t_2 - \tau}{t_2 - t_1} x$$

is continuous $\forall \tau \in [t_1, t_2]$ and any x on which h_A is defined. Hence, η_A satisfies the continuity conditions for homotopy.
Proof of Theorem 4: i)Let $V(x,t) = e^{-\delta t}\Omega(x,t)$. First we will show that V is continuous before proving the brunt of the theorem in ii). Since $e^-\delta t$ is continuous $\forall t$, it will be sufficient to check the continuity of Ω. It will be convenient to itemize the continuity conditions as follows:

1. Since $\Omega(A, t)$ is constant on $[0, t_1]$ and also on $[t_2, \infty)$ and $\Omega(B, t)$ and $\Omega(C, t)$ are constant on $[t_1, \infty)$, Ω is continuous on these intervals for each value.

2. As proven in Theorem 2, η_B is continuous on $B \times [0, t_1]$, and η_A is continuous on $A \times [t_1, t_2]$.

The last condition we must check are for the endpoints of each interval. However, it is easy to see that

$$\lim_{t \to t_1^+} \Omega(B, t) = \lim_{t \to t_1^+} U(\eta_B^t(B)) = U(\lim_{t \to t_1^+} \frac{t_1 - t}{t_1} h(B) + \frac{t}{t_1} B) = U(B)$$

$$= \lim_{t\to t_1^-} U(B) = \lim_{t\to t_1^-} \Omega(B,t)$$

Similarly,

$$\lim_{t\to t_2^-} \Omega(A,t) = \lim_{t\to t_2^-} U(\eta_A^t(A)) = U(\lim_{t\to t_2^-} \frac{t-t_1}{t_2-t_1} h_A(A) + \frac{t_2-t}{t_2-t_1} A) = U(h_A(A)$$

$$= \lim_{t\to t_2^+} U(h_A(A)) = \lim_{t\to t_2^+} \Omega(B,t)$$

Since we have examined the conditions for each point in $\{A,B,C\}$, we conclude that $\Omega(x,t)$ is continuous on $\{A,B,C\} \times [0,\infty)$.

ii) First, note that $U(h_A(A)) = U(\eta_A^{t_2}(A))$. Then $\forall t > t_2$,

$$e^{-\delta t_2} > e^{-\delta t} \Rightarrow e^{-\delta t_2}\Omega(A,t_2) > e^{-\delta t}\Omega(A,t) \Rightarrow V(A,t_2) > V(A,t)$$

It is clear that this is also true for bundles B and C. So a maximum most occur on $[0,t_2]$. Additionally, $[0,t_2] = [0,t_1] \cup [t_1,t_2]$. So, we can examine the maximum on each of these compact subintervals, and take the maximum of these extremal values. The existence of each is guaranteed by the compactness of the intervals and the continuity of V. Recall that by assumption, $V(C,t) < V(x,t)$ where $x \in \{A,B\}$ and $\forall t$. We will first examine $[0,t_1]$. $\forall t \in [0,t_1]$ such that $t \neq 0$,

$$e^0 = 1 > e^{-\delta t} \Rightarrow V(B,0) > V(B,t) \text{ AND } V(A,0) > V(A,t)$$

If $V(A,0) \geq V(B,0)$,

$$\Rightarrow V(A,0) > V(A,t) \text{ AND } V(A,0) > V(B,t)$$

If $V(B,0) \geq V(A,0)$,

$$\Rightarrow V(B,0) > V(A,t) \text{ AND } V(B,0) > V(B,t)$$

Hence, the maximum on $[0,t_1]$ is an element of $\{V(A,0), V(B,0)\}$.

We now move to $[t_1,t_2]$. By an argument symmetric to the one above, we have $V(B,t_1) > V(B,t)$ $\forall t > t_1$. Additionally, the function $V(A,t)$ is continuous on $[t_1,t_2]$. Thus, $\exists \tau \in [t_1,t_2]$ such that $V(A,\tau) \geq V(A,t)$ $\forall t \in [t_1,t_2]$. If $V(B,t_1) \geq V(A,\tau)$

$$\Rightarrow V(B,0) > V(B,t_1) \geq V(A,\tau)$$

and therefore the maximum on $[0,t_2]$ is an element of $\{V(A,0), V(B,0)\}$. If $V(A,\tau) \geq V(B,t_1)$, then $V(X,t)$ attains its maximum on $[t_1,t_2]$ at $V(A,\tau)$. Then

$$\max_{x\in\{A,B.C\},t\in[0,\infty)} V(x,t) = \max\{\max_{x\in\{A,B.C\},t\in[0,t_1]} V(x,t), \max_{x\in\{A,B.C\},t\in[t_1,t_2]} V(x,t)\}$$

$$= \max\{V(A,\tau), \max\{V(A,0), V(B,0)\}\}$$

which is plainly an element of $\{V(A,\tau), V(A,0), V(B,0)\}$.

References

[1] **de Clippel, Geoffrey and Kfir Eliaz,** "Reason-Based Choice: A Bargaining Rational for the Attraction and Compromise Effects," *Theoretical Economics*, forthcoming

[2] **Debreu, Gerard,** "Review: Individual Choice Behavior," *The American Economic Review*, 1960, *50* (1), 186-188.

[3] **Gul, Faruk, Paulo Natenzon, and Wolfgang Pesendorfer,** "Random Choice as Behavioral Optimization," Working Paper, Princeton University, February 2012

[4] **Huber, J., J.W. Payne, and C.Puto,** "Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis," *Journal of Consumer Research*, 1982, *9* (1), 90-98

[5] **Huber, Joel and Christopher Puto,** "Market Boundaries and Product Choice: Illustrating Attraction and Substitution Effects," *The Journal of Consumer Research*, 1983, *10* (1), 31-44.

[6] **Luce, R. Duncan,** *Individual Choice Behavior: a Theoretical Analysis,* Wiley New York, 1959

[7] **Natenzon, Paulo** "Random Choice and Learning," Working Paper, Princeton University, December 2010

[8] **Ok, E.A., P. Ortoleva, and G. Riella,** "Revealed (P) Reference Theory, Working Paper, August 2011

[9] **Shafir, Eldar, Itamar Simonson, and Amos Tversky,** "Reason-based Choice," *Cognition*, *49*, 11-36

[10] **Tversky, A.,** "Choice by elimination," *Journal of mathematical psychology,* 1972, *9* (4), 341-367

-, "Elimination by aspects: A theory of choice," *Psychological Review*, 1972, *79* (4), 281-299.

[11] **Tversky, Amos and Eldar Shafir,** "Choice Under conflict: The Dynamics of Deferred Decision," *Psychological Science*, *3* (6), 358-61