Discovered Preferences for Risky and Non-Risky Goods

Jason Delaney
Department of Economics & Finance
University of Arkansas at Little Rock
jjdelaney@ualr.edu

Sarah Jacobson[1]
Department of Economics
Williams College
Sarah.A.Jacobson@williams.edu

Thorsten Moenig
Department of Risk Management and Insurance
J. Mack Robinson College of Business
Georgia State University
thorsten@gsu.edu

March 2012

Abstract:
Are preferences stable or do they evolve with experience? While the assumption of stable preferences seems reasonable for many items, the stochastic nature of experience with risk could make risk preferences appear unstable because of incomplete learning. This accords with evidence of apparent instability of or evolution in risk preferences from lottery experiments. We develop a model of preference learning that could yield both well-formed (fully learned) preferences for most non-stochastic goods and imperfectly-formed preferences for stochastic items. In the model, an agent's value for a non-stochastic good is learned with a single experience but her value for a stochastic good requires several experiences to be learned. When infinite time has elapsed, nearly all stochastic good and stochastic goods have their values fully updated; however, because of this difference, at finite time stochastic goods are less likely to be correctly valued. Further, if learned preferences are imperfectly remembered (if they decay), agents' values for stochastic items tend to remain farther from their true values as compared to values for non-stochastic items. This model retains stable inherent preferences, but allows for evolution of expressed preferences in a predictable and intuitively appealing manner.

JEL codes: D81, D83, D01, D03, C91
Keywords: risk preferences, learning, discovered preferences, preference stability

---

[1] Corresponding author. This work is a very early draft and comments are welcome.

Introduction

Are preferences stable? Do people know their own preferences? Any economic analysis that depends on a utility function or revealed preferences implicitly assumes that the answers to these two questions are usually, "Yes!" and "Yes!" However, there is reason to believe that in the case of risky (stochastic) experiences, normal adults can at equilibrium have imperfectly learned preferences. These imperfectly learned preferences can look like unstable preferences. It seems plausible that most normal adults know their own preferences over non-stochastic goods or experiences fairly well: when you taste a particular ice cream flavor you get immediate feedback as to whether you like it. However, failure to perfectly know one's own preferences over risk is plausible: when you experience a random event the affect experience is far less direct and after-the-fact introspection may confuse the risky good with its *ex post* value. In this paper, we use a model of learning of preferences to explore some situations in which people may rationally hold imperfectly-learned conceptions of their own preferences for risk.

Risk preferences are a fundamental part of modeling individual microeconomic behavior. Our understanding of those preferences continues to evolve as we develop new theories and test them in the lab. In the process of this testing, attention has been paid to behavioral biases and subject error. Behavioral biases may represent true preferences, but subject error can interfere with the expression of true risk preferences. Many experimental economists have observed subjects' choices over repeated risky choice evolving in ways that are inconsistent with stable preferences over risk. While some of this must represent classic subject error, we hope to contribute to the discussion of mis-represented preferences by recasting part of this error as a failure to know one's own preferences. If a person does not know her own risk preferences, we may be able to observe the by which process she learns those preferences when she makes repeated choices over risk. The learning process is inherently interesting for the purposes of predicting or understanding human choice, and it has welfare implications if it implies people at equilibrium choose in ways that violate their preferences. Further, learning of preferences may also be conflated with other elements of interest to researchers.

Psychologists sometimes argue that preferences are not stable but are instead constructive (e.g., Fischhoff et al., 1988). This is unappealing to economists because analysis built from the assumption of stable preferences is powerful and often describes reality well. Arguments for preference stability are best demonstrated in Stigler and Becker (1977): it is better, they argued, to model changes in choice over time as changes in constraints (including building of specific capital that convert particular goods into utility) rather than changes in preferences.

Rather than perfectly unstable or perfectly stable preferences, some have considered the possibility that people have preferences but that they must discover them. The idea of discovered preferences was articulated in Plott (1996), who proposed that if discovered preferences is a meaningful phenomenon, then it implies that people always try to optimize but at first don't know enough to be successful at optimization. Then repeated choice with feedback allows people to arrive at consistent and stable preferences.

In a discussion of the discovered preference hypothesis, Braga and Starmer (2005) draw a distinction between "institutional learning," or learning the structure of the institution in which they are making choices, and "value learning," or learning of one's own preferences. In this paper, we focus particularly on value learning. Cubitt et al. (2001) suggest that discovered preferences be interpreted in the context of risky choice with the idea that there is a true relationship between an experience and a person's affect relating to that experience, but that people may need to learn about that relationship by repeated consumption.[2]

Preference discovery is at odds with the Stigler and Becker (1977) approach; Stigler and Becker would prefer that such learning be represented as changes in constraints. As argued in Andersen et al. (2008), however, the distinction between stationary but state-contingent preferences and preferences that can change is a very fine one.

Kahneman and Snell (1990) note that when experiences are familiar and immediate, people seem fairly good at predicting the utility they expect to get from the experience. This also relates importantly to the distinction between "decision utility," the value system that drives individuals' decisions, and "experienced utility," the utility people feel experiencing what they consume or the utility they remember feeling (Kahneman et al., 1997). For risky goods, experienced utility is hard to assess because the item being valued is fundamentally the *ex ante* gamble, while the experience includes both the *ex ante* gamble and the *ex post* realization.

There is experimental evidence of evolution of or instability in expressed preferences in non-stochastic domains. In many of these studies, it's difficult to distinguish "institutional learning" from "value learning". Studies showing convergence to a value that is known to be true because it is induced (e.g., Noussair et al., 2004) demonstrate institutional learning. Other studies find that errors or biases decline with repeated choice. These may also indicate institutional learning, including studies on the WTA-WTP gap (e.g., Coursey et al., 1987; List, 2003; Shogren et al., 2001; Shogren et al., 1994) and non-dominant bidding behavior (List and Shogren, 1999).[3]

[2] The role of affect in the learning of preferences is distinct from the role of affect in forming preferences. It is the latter that is discussed in affect literature like Isen (Isen, 2005), who show that moods affect choice behavior. This is a question of exogenous moods changing apparent ranking, not about an agent learning her own true ranking.

[3] However, Knetsch et al. (Knetsch et al., 2001) show that in an auction may not decrease but actually increase the WTA-WTP gap.

Discovered preferences has received attention in the realm of environmental valuation, since items being valued are often unfamiliar or are items that respondents have never considered valuing before. Several valuation studies find that errors appear to decline with repeated trials: Kingsley and Brown (2010) find that intransitive choice, preference reversals, and estimated error decline (although apparent preference remains stable); Ladenburg and Olsen (2008) find that starting point bias declines; and Bateman et al. (2008) find that the gap between values elicited through single-bounded and double-bounded procedures declines. In these environmental valuation studies, the respondents appear to be refining their choices; is this institutional learning or value learning? The respondent is not exactly engaging in repeated consumption, which should be the key to preference discovery. Either these represent institutional learning, or repeated introspection is sufficient to familiarize a person with her own preferences.

There is also significant evidence of expressed preference for risk changing. The process of learning (either institutional or value) should be different if lotteries are realized and feedback is given throughout the process; without realizations and feedback, subjects' only chance to learn is by introspection. Over repeated trials without feedback, Cox and Grether show that preference reversals decline (Cox and Grether, 1996), although Braga et al. (Braga et al., 2009) show that further repetition may cause other anomalies. Several studies of repeated lottery experience without feedback (Birnbaum and Schmidt, 2009; Hey, 2001; Loomes et al., 2002) find that estimated error rates or choice inconsistencies appear to decline with repetition. Keren and Wagenaar (1987) and Loomes Sugden (1998) find a decrease in expected utility violations, while Bone et al. (1999) finds an increase therein, with repetition.

The studies that examine repeated lottery exposure with feedback are limited, and often focus on situations in which the agent is learning probabilities. Thaler et al. (1997) find that more frequent feedback causes repeated risky choice to become more risk averse. Barron and Erev (2003) and Hertwig et al. (2004) find that subjects who are told lottery odds perform differently than those who learn about the lotteries through experience with feedback. Here there is obvious institutional learning that must occur, so it is impossible to disentangle discovery of preferences. This sort of learning of the value of uncertain prospects is modeled and simulated in March (1996). Jessup et al. (2008) describe probabilities to subjects and find that subjects who get feedback move toward choosing according to the objective probabilities while without feedback they overweight small probabilities. Van de Kuilen and Wakker (2006) find that repeated trials without feedback don't reduce Allais violations but with feedback the violations do decrease.

What do we know about risk preference stability over a longer period of time? If life experience in general allows agents to learn about their risk preferences (or otherwise allows risk preferences to change), then observed preferences may appear unstable over time. Horowitz (1992) finds individual temporal instability but aggregate temporal stability over six weeks, but Harrison et al. (2005) find a fair amount of individual temporal stability over four weeks. Andersen et al. (2008) look at risk preference stability using lottery choices by Danish adults across 17 months and find some variation but a general stability, as well as some sensitivity to the current financial state of the respondent. Still, it is unclear what experiences the subjects had to learn from in the intervening periods, and it is unclear how much spillover there is between experience with one risky asset and learning one's preference for a wholly different risky asset.

If we are concerned with learning of preferences (values), we can look to existing models of learning in other contexts. Thorndike (1898) inspired a school of thought based on the "law of effect:" the prior you have about how you should behave (or how much you like something) is updated based on your experienced outcomes. Models of reinforcement learning (Bush and

Mosteller, 1955) or fictitious play (Brown, 1951) have been developed, though mostly they have been used to model learning of unknown probabilities or dynamics of strategic interactions. Sarin and Vahid (1999) lay out a simple model of learning based on updating valuations with experienced outcomes; our model will be similar but when applied to risky assets will explicitly consider the *ex ante* nature of a lottery. Cohen et al. (2008) explore how experience may color risk perception, making agents more or less pessimistic, which they represent as a weighting function of the probabilities. This is not a model of learning but of a shifting of preferences, since the value function actually changes given known probabilities. We are interested in a similar approach, but within a framework of refining preferences toward a true inherent risk preference. Refinements of dynamic probabilistic choice such as decision field theory (1993) also inform our understanding of the cognitive processes of learning and reflection about the values of alternatives.

We hope to add to the literature on risk preferences by contributing a model of preference discovery that is consistent with the methods and mindset of neoclassical economics but that could yield the appearance of unstable risk preferences. This model will allow the idea of discovered risk preferences to be tested carefully. If preference learning is important in the context of risk, this has implications for researchers and also may suggest the presence of real welfare losses that could be avoided.

In the following sections, we first develop a simple model of preference learning, showing that normal adults should have fully stable preferences over most common goods if learning does not decay and goods are non-stochastic. Next, we extend the model to a case in which the consumption good yields stochastic outcome for the consumer. We show that as long as experience occurs enough times, preferences for the risky good become and remain stable. We explore two situations in which we show that preference learning is subverted: a case in which the consumer can choose to avoid future experience with the risky asset, and a case in which learning decays between consumption experiences. In the former case, we show that a person may consistently express apparent preferences different from her true preferences. In the latter case, we show that preferences may appear unstable. In both cases, repeated experience would force expressed preferences to approach the person's true preferences. In the final section, we conclude with implications of the model and we discuss how it can be tested.

Setup

Let there be a set of *K* possible alternative strategies $X = \{x_1, x_2, \ldots, x_K\}$. These strategies represent choices of prospects that may be amounts of money, simple consumption goods, or stochastic experiences (things whose outcomes are determined by some random process). At particular moments in time *t*, the agent faces a choice set: a subset of k_t exhaustive and mutually exclusive strategies $X_t = \{x_1, x_2, \ldots, x_{kt}\}$. At time *t*, she has some value $V_t(x_i)$ for each strategy x_i; this is the satisfaction she associates with it at time *t*.

We assume that a person, faced with alternatives, chooses the option that she expects to yield her the highest possible satisfaction or utility of all available strategies. That is, she will choose alternative x_i if and only if $V_t(x_i) \geq V_t(x_j)\ \forall\ j \neq i$.

We could instead write the agent's value for a strategy as $V_t(x_i \mid v_0)$, where v_0 refers to the level of satisfaction she had before consuming the good. For simplicity and ease of expression we omit that argument. This can be done without loss of generality if present consumption is separable from status quo wellbeing or if people can calibrate their expected satisfaction from a consumption experience to their starting level of wellbeing. If either of these does not hold, this omission is problematic. We return to this point in the conclusion.

Preference Learning for Non-Stochastic Goods

Let the choice alternatives facing the deciding agent be non-stochastic consumption items, and let her satisfaction from each also be deterministic. In this case, choice of strategy x_i will provide her the deterministic satisfaction $\tilde{V}(x_i)$ with probability 1. For example, she may be facing a basket of apples, oranges, pears, and bananas. Each apple is of a consistent quality and yields the same consumption experience (and this is also true of the other fruit).[4]

Does the agent necessarily know how much satisfaction each option will provide her? If she has undertaken a strategy before, she most likely remembers how she liked the experience. For example, for food, there is evidence that people learn and retain their tastes for significant consumption items; see Rozin (1982) and Rozin and Vollmecke (1986). On the other hand, if she has not previously consumed a specific good, she may not have a perfect prediction of how it will please her. She instead has a prior belief $V_0(x_i)$ as to each strategy's value. This prior value may be informed by an assessment of the characteristics of the outcome the strategy will provide or by comparisons to similar strategies that she has tried.

It is also possible to model this prior as a distribution of possible values rather than a point value for each strategy. Our model with a degenerate prior is essentially a model of "preference correction" from an incorrect prior, while a model with a diffuse prior would demonstrate "preference refinement" as a broad distribution is narrowed toward a point estimate. We leave this "preference refinement" model for future work. However, unless the agent is

[4] Obviously, in reality, all apples are stochastic experiences.

ambiguity averse, she may treat a diffuse prior exactly as she would treat a degenerate prior with the same expected value if she behaves as if her value was the mean of the prior distribution.

If at time *t* a person has never chosen a given strategy, her value for it may still be her prior $V_t(x_i) = V_0(x_i)$. Alternatively, if her values are informed by experience with other strategies, she may update the prior to reflect her new understanding; for simplicity we exclude that possibility.

If a person chooses strategy x_i, she experiences value $\tilde{V}(x_i)$. If she knows that x_i is deterministic in value, it should be the case that whatever her prior belief $V_t(x_i)$ about the strategy's value to her, her posterior belief after consuming it should be $V_{t+1}(x_i) = \tilde{V}(x_i)$. Once updated to that true value, as long as her learned preference does not decay, she will retain her true value as her believed value. Let us define a history vector $\mathbf{h}_t$ to contain all of the strategy she has ever chosen as of time *t*. We must now condition her predicted value now on both the good and her consumption history so that $V_t = V_t(x_i, \mathbf{h}_t)$. Therefore, for all previously-chosen strategies $x_i \in \mathbf{h}_t$ it must be the case that $V_t(x_i, \mathbf{h}_t) = \tilde{V}(x_i)$.

Proposition 1a (full learning on non-stochastic outcomes with forced experience): Suppose all strategies appear in choice sets with positive probability and each choice set contains one or more strategies. Then as time approaches infinity, for all strategies, the agent's value for the strategy is updated to her true value with probability 1: as $t \to \infty$ $V_t(x_i) \to \tilde{V}(x_i) \; \forall \, i$.

Proof: If a strategy appears in any given time with positive probability, then it must be chosen eventually since as time approaches infinity she must have faced a choice set with only this strategy available at least once. For all times after time t_i in which she first chose the strategy, it must be the case that $V_t(x_i, \mathbf{h}_t) = \tilde{V}(x_i)$ because after t_i, $x_i \in \mathbf{h}_t$.

Proposition 1b (nearly-full learning on non-stochastic outcomes with pairwise selection): Suppose all strategies appear in choice sets with positive probability and each choice set contains two or more strategies. Then as time approaches infinity, for all strategies except the lowest-ranked strategy x_w *such that* $V_0(x_w) = \min\{V_t(x_1, \mathbf{h}_t), \ldots, V_t(x_K, \mathbf{h}_t)\}$*, the agent's value for the strategy is updated to her true value with probability 1: : as* $t \to \infty$ $V_t(x_i) \to \tilde{V}(x_i) \; \forall \, i$.

Proof: In each choice set, the agent chooses the strategy she believes to be highest in value. If the agent is never forced to experience a strategy by virtue of it being her only option, then, by the same logic as the proof for 1a, all strategies will eventually be chosen except for any strategies that are lowest-ranked in every choice set in which they appear. If choice sets get as small as two alternatives, as time approaches infinity, then every possible pair of strategies will appear. Only the lowest-ranked strategy will be the worst strategy in every choice set in which it arises.

Note that because values for strategies can be updated both upward and downward, there is no guarantee that the lowest-ranked strategy at any given time is the strategy with the lowest true value. By the same token, the lowest-ranked strategy may be initially overvalued and thus may be chosen at some point; if the update moves this strategy value below the believed value of all other strategies at that moment, then as time approaches infinity all strategies will be chosen and thus all strategies will reach their true value.

These propositions state that all (or nearly all) strategies will, given enough time, be updated to their true values. Relatedly, let us define a "common" good as a good for which the strategy of choosing it appears frequently in choice sets. Therefore, as long as the prior value for a common strategy is not very low, normal adults with a reasonable amount of life experience (i.e. for whom t is finite but large) have correct valuations for these strategies with high probability. So if apples appear in enough fruit baskets, unless we have an exceedingly low prior value for apples then we should have all tried apples by the time we are adults, and thus as adults we should have a correct value for apples.

Note that the agent is somewhat naïve here, in that she doesn't literally know when her value for a strategy is incorrect. However, she does remember what strategies she has and has not tried, and in this model she can assume that if she has not tried a strategy her value for it is incorrect. But she does not treat a known-incorrect value differently from a known-correct value: she still chooses the (apparently) highest-ranked item available in each choice set.

Preference Learning for Stochastic Goods

Let y_i be a strategy yielding a stochastic outcome with probabilities known to the agent, a probability distribution over N possible outcomes z: $y_i = \left(z_1, z_2, \ldots, z_N; p_1, p_2, \ldots, p_N\right)$. Let these outcomes all be deterministic in value. Assume that the probabilities and potential outcomes are fully known to the agent. Further, let the outcomes be common goods that are not dominated by all other goods: in other words, assume that at time t the agent can correctly value each outcome if it occurs deterministically: $V_t\left(z_i\right) = \tilde{V}\left(z_i\right) \forall z_i$. These strategies can be thought of as lotteries over money outcomes, where we assume the agent knows how much she values the money amounts she may win or lose. Alternatively, we can imagine that on each choice occasion the agent must pick a restaurant y_i for dinner. Each restaurant provides a random quality because the staffs vary. The agent knows her value for all possible meals z_j within the range of meal qualities at each restaurant and she knows how likely p_j each meal quality is at each restaurant.

As with non-stochastic strategies, the agent starts with a prior belief as to her value for each stochastic item $V_0(y_i)$ and has a true value for that stochastic item $\tilde{V}(y_i)$. These values may yield rankings that may reflect any model of choice under risk, including expected utility, prospect theory and its variants, or rank dependent utility. For the purpose of preference learning, we simply treat the agent's values, however, generated, as values.

How does preference learning work for stochastic items? If the agent consumes a stochastic item, the experience is not as direct as the feeling of consuming an item that is deterministic. That is, if she consumes stochastic item y_i, she does not literally experience value $\tilde{V}(y_i)$. This is because an agent's value for a stochastic item (her prior or her true value) is necessarily an *ex ante* value: it only has meaning before the lottery has been realized, that is, before the uncertainty has been resolved. However, the value she experiences if she chooses strategy y_i at time *t* is first her prior value $V_t(y_i)$, some instantaneous feeling of being about to face a lottery, and then the outcome that is realized (some z_j). None of these is exactly her *ex ante* value. It is possible to think of this amalgam of feelings as a temporally extended outcome, as in Kahneman (1997), who demonstrates that when outcomes are extended, remembered utility seems to be an average of the peak of the experience and the sensations experienced at the end.

Regardless, when she implements a strategy (chooses a restaurant), she will experience some level of satisfaction $v_t(y_i, z_t)$ based on the lottery experience and the realized outcome (the feeling of anticipation and the meal she ends up being served). She will not directly replace her old value $V_t(y_i)$ because she knows this is a stochastic process and individual realizations do not represent her true value. However, she may realize that the experience of facing the lottery felt either better or worse than she had expected. By the same token, it may not necessarily be the case that a high realization (good meal) causes her to update her value upward or that a low realization (bad meal) causes her to update her value downward. If she finds that the pleasure of a good meal doesn't compensate as much as she'd expected for the anxiety of uncertainty, she may update her value downward after a high realization, for example.

What structure can we add to the updating process? A reasonable updating process meets three criteria. First, some memory of previous experiences must be retained. Second, as the agent has experienced the strategy enough times that her observed frequency distribution of the outcomes approaches the objective probability distribution, her value should approach the true value. The latter criterion is based on the point that a stochastic experience is a probability distribution across outcomes; if the experience is difficult for agents to value because of its randomness, then experiencing that whole distribution should allow agents to learn their values. Without such a criterion, there would not necessarily be convergence of the process. Without convergence, the agent would not have stable preferences for risk in a meaningful way. Third, while the agent's value can be updated in the wrong direction on any given occasion, it should be the case that updates on average move the agent's believed value toward her true value.

We choose a fractional updating method using learning parameter λ as in Sarin and Vahid (1999). Suppose strategy y_i is chosen at time *t* and outcome z_t is realized, and this is the n^{th} time that the agent has experienced the strategy. Let p_{min} be the least probability of all in the probability distribution for strategy y_i ($p_{min} = \min\{p_1, \ldots, p_m\}$ if y_i has *m* possible outcomes). Then the agent updates her value for this strategy as follows:

$$V_{t+1}(y_i) = (1-\lambda)V_t(y_i) + \lambda\phi\left(v_t(y_i, z_t), \tilde{V}(y_i), n\right)$$

Let γ be a positive integer representing the number of times the distribution should be experienced for full learning. Then we require that as $n \to \frac{\gamma}{p_{min}}$, $\phi\left(v_t\left(y_i, z_t\right), \tilde{V}\left(y_i\right), n\right) \to \tilde{V}\left(y_i\right)$. If this is the case, then after trying the strategy $\frac{\gamma}{p_{min}}$ times, the agent's value for the strategy stabilizes at its true value (since thereafter it will be updated with its true value). Alternatively, we could say that as $n \to \infty$, $\phi\left(v_t\left(y_i, z_t\right), \tilde{V}\left(y_i\right), n\right) \to \tilde{V}\left(y_i\right)$. However, this does not demonstrate the important relationship that should exist between the number of trials of a strategy, the probability distribution, and the agent's ability to learn. The relationship we propose implies the intuitive result that strategies whose stochastic outcomes exhibit a higher variance will show a slower rate of value learning. Note that if the stochastic outcomes are degenerate so that $p_{min} = 1$, this implies full learning as $n \to \gamma$. For our non-stochastic outcome case in the previous section, we simply assumed $\gamma = 1$, which is reasonable if an agent knows the outcome is not stochastic and so knows she has fully experienced the outcome after experiencing it once.

Proposition 2a (full learning on stochastic outcomes with forced experience): Suppose all strategies appear in choice sets with positive probability and each choice set contains one or more strategies. Then as time approaches infinity, for all strategies, the agent's value for the strategy is updated to her true value with probability 1: as $t \to \infty$ $V_t\left(y_i\right) \to \tilde{V}\left(y_i\right) \forall i$.

Proof: Although values for strategies with stochastic outcomes are not updated fully on their first trial, the logic is identical to the logic for Proposition 1a: if time is infinite and the agent is occasionally forced to choose each strategy on multiple occasions, eventually she will fully update her value for each strategy.

Proposition 2b (nearly-full learning on stochastic outcomes with pairwise selection): Suppose all strategies appear in choice sets with positive probability and each choice set contains two or more strategies. Then as time approaches infinity, for all strategies except the lowest-ranked strategy y_w *such that* $V_t(y_w) = \min\{V_t(y_1), \ldots, V_t(y_K)\}$*, the agent's value for the strategy is updated to her true value with probability 1: as* $t \to \infty$ $V_t(y_i) \to \tilde{V}(y_i) \, \forall \, i$.

Proof: The proof follows the same lines. Note that in this case, a strategy that has been tried can be incorrectly devalued to become the lowest-ranked strategy, derailing its updating process. Thus a strategy that has been tried may never be fully updated to its true value.

If the agent is forced to attend each restaurant occasionally because of a limited choice set, she will eventually converge to her true value for each restaurant. If she always has at least two restaurants open to choose from, then she may retain an incorrect value for the one she least prefers but she will find her true values for all of the others.

What might be observed in finite time? If elapsed time is not large enough in comparison to the frequency with which a given strategy y_i is in the choice set to create a high likelihood that the strategy has been chosen multiple times, an agent may retain an incorrect value for that strategy. This is essentially an off-equilibrium case unless the set of possible strategies is arbitrarily large and/or the probability that this strategy appears in a choice set is arbitrarily small. Still, this off-equilibrium scenario is more likely in the case of a strategy with stochastic outcomes than in the case of non-stochastic outcomes, since in the latter case a single experience allows an agent to fully learn her value for the strategy.

Proposition 3 (slower learning for stochastic outcomes): at finite time t, the agent is more likely to have learned her true value for a strategy with a non-stochastic outcome as compared to a *strategy with stochastic outcomes given similar common-ness (likelihood of appearing in choice sets) and ranking in the set of all possible strategies.*

Proof: If the strategies have similar ranking and similar common-ness, their first trial will occur on expectation at the same time. Multiple trials are required to learn the value of a strategy with a stochastic outcome while only one is required to learn the value of a strategy with a deterministic outcome. Therefore, more time must elapse before the value of the strategy with stochastic outcomes is fully learned.

This must be true since on expectation, strategies with similarly frequent appearance in choice sets and similar rankings have been tried the same number of times. If a strategy with non-stochastic outcomes is tried once, its value is perfectly learned; it takes repeated experience to learn a value for a strategy with a stochastic outcome. By the same argument, any strategy that takes more retrials before the agent learns her true value for it will take longer time to learn. As a result, strategies with more outcomes and strategies with longer odds will take longer to learn.

Additionally, the less common a strategy, the less likely it is that the agent has learned her true value. Additionally, the lower-ranked a strategy is, the less likely it is that she has learned its true value. In the limit, of course, the lowest-ranked strategy's value may never be fully updated. If she underestimates the value of this strategy by a large margin, she could rank it incorrectly relative to her other options. In our analogy, she could be undervaluing a restaurant experience that she would prefer to others simply because she refuses to sample it. Her wellbeing could be increased if she were forced to visit this restaurant. This kind of misapprehension is possible for strategies with both non-stochastic and stochastic outcomes but is more likely for strategies with stochastic outcomes because even a strategy that has been experienced can (as result of a negative update) be incorrectly demoted to a position at the lowest rank.

The agent would be willing to have a lowest-ranking alternative removed from her choice sets if such a removal were costless. If retaining options is costly, she may opt to remove her lowest-ranking alternative and possibly others at the bottom of her ranks. (This is one case in which a diffuse prior could yield a different prediction as compared to a degenerate prior.) If there were some chance that she would be forced to choose her lowest-ranked strategy, she would be willing to pay to remove that strategy from her future choice sets.

Decay of Learning

All of the above assumed that once a value is learned, it is only updated with experience and the updated value remains known to the agent in future time periods. In reality, it may be the case that once a value has been learned, the agent does not perfectly retain what she has learned. It may be that as time elapses between experiencing the strategy, her value for it changes. That is, it may be that we should express the agent's value for strategy y_i at time *t* as:

$$V_t(y_i) = V_{t-1}(y_i) + \delta(y_i, t, t_{last})$$

Here, δ represents a learning decay function, and t_{last} is the last time the agent tried the strategy. What criteria should such a decay function meet? First, the distance between the last-updated value and the currently-held value must be non-decreasing with time. Second, the decay should occur in the direction of the agent's original prior value for the strategy (toward $V_0(y_i)$). Third, the decayed value should converge to the original prior value as time approaches infinity. In other words, as time passes, the agent should move toward and eventually return to a state of ignorance about the strategy. Such a decay function could be exponential in form.

In our restaurant example, immediately after visiting a restaurant, the agent updates how she values it and how she ranks it relative to other restaurants. As the days pass before she returns there, her memory fades and she reverts her value slowly toward the original impression she had of the restaurant before visiting it. In fact, her memories of all of her restaurant choices are fading. They may be fading at different rates and in different directions as each value moves toward her original impression for that restaurant. It may be that her ranking across the restaurants changes even without any further experience, simply because of differential decay.

How will such decay affect the value that an agent holds for a strategy at any given time? In the case of a strategy with a deterministic outcome, when the strategy is tried it will be updated to its true value; thereafter it will decay toward the prior until the next trial. This will appear as an irregular oscillation of value whose frequency is determined by the rate at which the strategy is tried (in turn a function of how common the strategy is and how highly ranked it is) and whose depth is determined by the rate of decay. Of course, the depth of the oscillation is limited by the distance between the true and prior values. This process is depicted in Figure 1a.

How will such decay affect the learned value for a strategy with stochastic outcomes? The pattern of the agent's value path will be similar but far less clean because some of the learning updates will be in the wrong direction and because multiple trials are required to fully learn the true value. For a strategy that is tried often (because it is common and highly-ranked), its value will oscillate in the neighborhood "below" the agent's true value for the strategy. This is depicted in Figure 1b. For a strategy that is tried less often, the decay process could offset the learning process completely, as depicted in Figure 1c. For strategies tried infrequently (because they appear less frequently in choice sets or because they are low-ranked relative to other strategies), decay can fully offset learning so that average value remains close to the prior value.

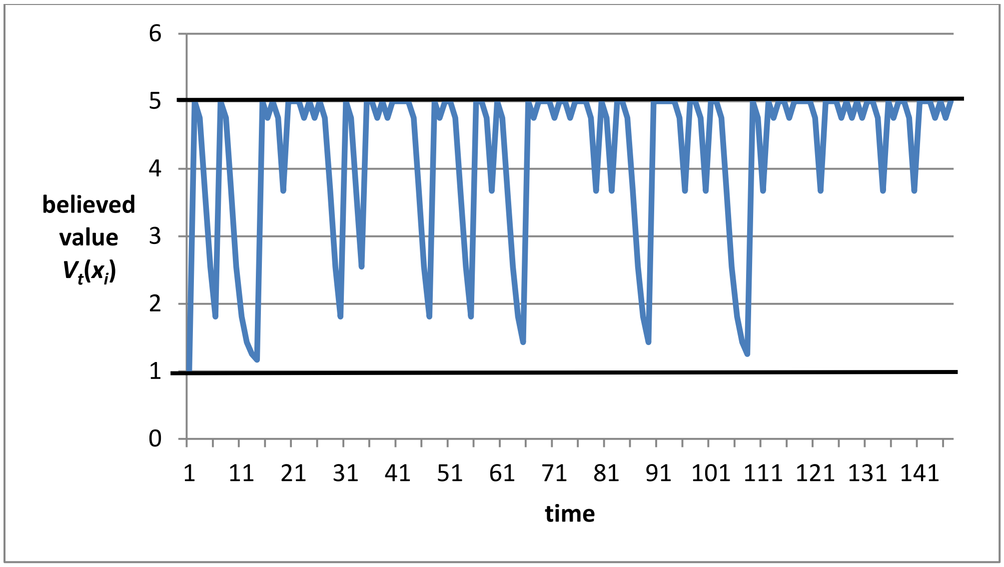

Figure 1a: Agent's believed value for a strategy with non-stochastic outcome

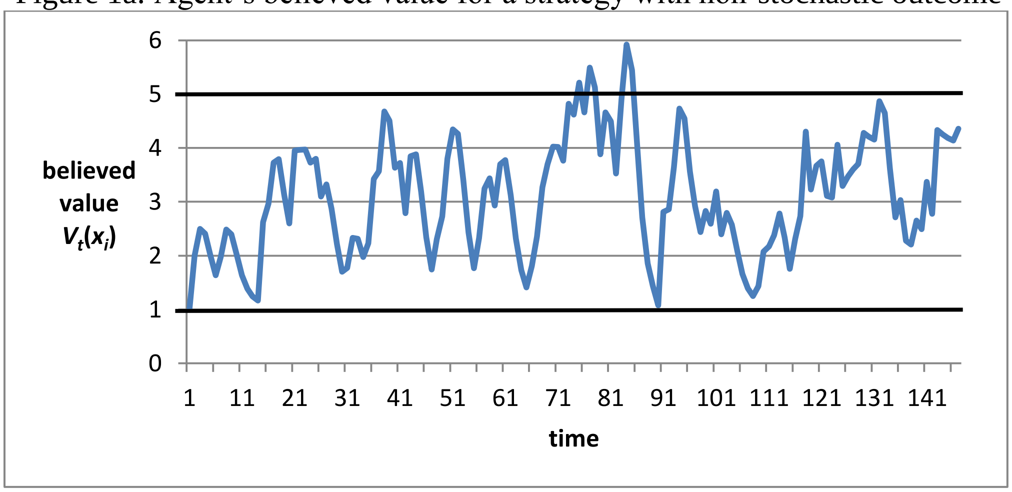

Figure 1b: Agent's believed value for a strategy with stochastic outcomes

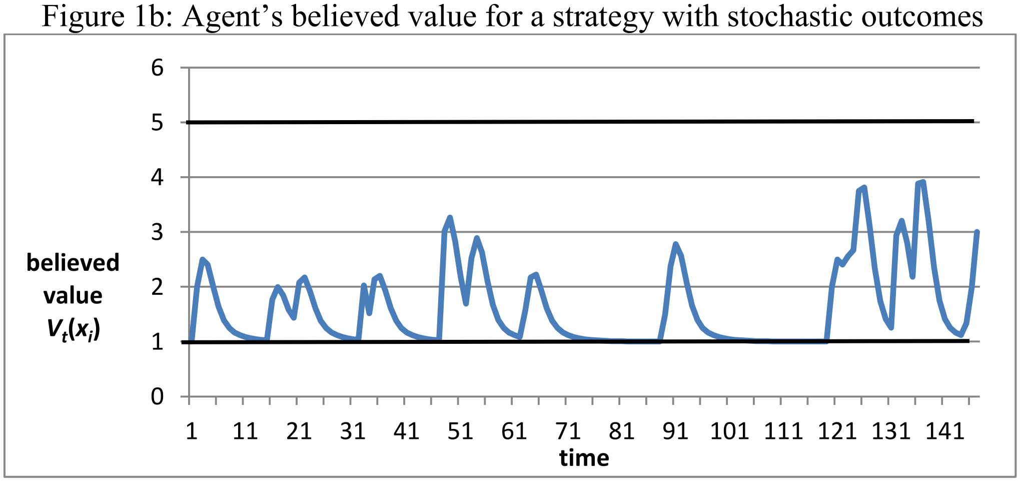

Figure 1c: Agent's believed value for a rarely-tried strategy with stochastic outcomes

Figure 1: Dynamics of agent's believed values over time with decay (with true value = 5, prior value = 1)

Proposition 4 (conditions favorable to preference learning in the presence of decay): If there is decay in learning, the time-averaged value held by an agent for a strategy will be closer to the agent's true value if the strategy is:

a. common as compared to if it is uncommon

b. high-ranked as compared to if it is low-ranked

c. yielding a non-stochastic outcome as compared to yielding stochastic outcomes.

Proof:

a. If a strategy is common, it appears relatively often in choice sets as compared to uncommon strategies. Therefore, all other things equal, it will be tried more often. Therefore its value will be more frequently updated to or toward its true value; therefore its value will spend more time near the true value.

b. If a strategy is high-ranked, it is chosen more often. By the same argument, its value will spend more time near the true value.

c. If a strategy yields a non-stochastic outcome, each trial updates the agent's value fully to the true value. If a strategy yields stochastic outcomes, successive trials partially update the agent's value toward her true value, while some may move in the wrong direction. Thus given a similar frequency of trial, all other things equal, the value of a strategy with a non-stochastic outcome spends more time near the true value.

Decay can interact with the learning process, particularly for strategies with stochastic outcomes. Since preference learning for strategies with stochastic outcomes is incomplete on a single trial, decay is more likely to offset learning. That is, if a strategy has stochastic outcomes, it is more likely to have an average value closer to the agent's prior value and farther from the agent's true value for the strategy (as compared to a strategy with non-stochastic outcomes that is similarly common and similarly ranked). It's also true, however, that a low-value strategy with stochastic outcomes can be incorrectly updated in a positive direction which may temporarily cause the agent to choose that strategy more often (if it moves past other strategies in the choice set). However, this is self-correcting, since each trial in expectation moves her value toward her true value.

Conclusion

We develop a model of preference learning to reflect the discovered preference hypothesis. We show that if preference learning occurs in this way, then at any finite time, an agent is less likely to have learned her true value for a strategy with stochastic outcomes than for a strategy with non-stochastic outcomes. We also show that this failure to complete the learning process is exacerbated if the agent can opt out of experience with a strategy and if learned preferences decay over time between experiences. The reason the learning process yields less complete results for strategies with stochastic outcomes than for strategies with non-stochastic outcomes is that the agent can learn her value for a strategy with a non-stochastic outcome in a single trial while it takes repeated experience for her to learn her value for a strategy with a stochastic outcome, and some steps in the learning process for a strategy with stochastic outcomes may actually move her believed value away from her true value.

If preference learning is a meaningful phenomenon, then agents in the real world make some decisions based on incorrect assessments of their values, particularly when stochastic outcomes are involved. However, if this is the case, it should be possible to detect the process of preference discovery using repeated experience; some existing evidence from the literature implies that some such learning may occur in laboratory experiments with repeated choice but in many of those it is difficult to disentangle institutional learning from value learning.

Further, if preference learning is important, individual welfare could at times be increased by forcing people to learn their own preferences—training with simulations of the situations of interest could help people learn their true values, although learning decay could offset some of these gains. At the same time, if preference learning is important, then researchers (particularly experimentalists) trying to elicit preferences may elicit incorrect preferences or may have their results contaminated by a process of preference learning if subjects undertake repeated tasks. This is also argued in Cubitt et al. (Cubitt et al., 2001), who employ one-task tests to demonstrate violations of expected utility theory that are robust to any discovery process.

Our model focused on a preference learning situation in which the agent does not know she has an incorrect value for the strategy; she believes with certainty that she holds a particular value but until she accrues experience and updates that believed value, she does not know her true value. An alternative specification would have the agent aware of her ignorance: she may know that she does not know her correct value for an as-yet-untried strategy. This could be represented as a diffuse, rather than degenerate, prior value, and learning would refine that distribution until it approaches a point. Unless the agent is ambiguity averse, the predictions of a model built from diffuse priors should be similar in most cases to the predictions of our current model. If the agent is ambiguity averse, she may tend to avoid trying new strategies simply because she prefers not to face a distributed value.

On a related note, our model deals with strategies whose outcomes are stochastic but are specified in terms of a known probability profile—that is, strategies whose outcomes are risky but not ambiguous (i.e., where Knightian uncertainty is absent). The case of Knightian uncertainty (ambiguity) is certainly of interest, since most situations faced by agents are best represented this way. However, if the agent does not know the probability distribution across outcomes, experience will grant her both value learning and institutional learning—learning of both her own value and the probability distribution. Note Thaler et al. (1997), Barron and Erev (2003), and Hertwig et al. (2004) look at this case. In fact, her own value can never be certain until she knows the probability distribution that applies. It is unclear how to represent the interaction between these two learning processes. This may be a fruitful area for future research.

Another essential issue is to what extent learning of the value of one strategy should spill over into knowledge of the value of another (as-yet-untried) strategy. If I eat an apple for the first time, does it teach me about my value for not just apples but pears? If I experience a coin-flip lottery for the first time, do I begin to learn my value for only that lottery—or for all lotteries? All coin-flip lotteries? All coin-flip lotteries which have outcomes on a similar scale? If such spillover learning can occur, it should reduce the chance that any strategy remains un-updated forever since all strategies have a higher chance of being updated.

A related question is how to consider these values within the framework of utility maximization. If the ranking implied by an agent's values can be represented by a utility function, there should be some coherent structure across values—if not for different goods (different arguments to the utility function) then certainly for different quantities of the same good. In particular, if we envision a utility function with money as an argument, the curvature of the utility function's projection describes risk preference in most economic theories of risky choice. Not knowing (or having an incorrect idea of) one's risk preference would imply not knowing the curvature of one's utility function. Thus incorrect valuation of a lottery with money outcomes would seem to imply incorrect valuation of deterministic money outcomes. How could a person know how to value a “sure thing” amount of money but have an incorrect value for a lottery?

It could be that she has separate utility functions: one for certain outcomes, and one for uncertainty. Andreoni and Sprenger (2010) explore this idea and find some experimental evidence supporting this approach. On the other hand, it could be that an agent builds her conception of her valuation function from a series of local approximations. In this case, the agent may not be aware of coherence across different kinds of goods. This idea is effectively rejects the idea of stable preferences in favor of something much more like constructive preferences, and seems unjustified since there is evidence that preferences are not wholly unstable. If we believe that an agent is coherent enough to have a utility function structuring her choice, we prefer to believe that she is not so naïve as to fail to recognize that coherence. It is by this token that we assume that we need not be concerned with the agent’s starting level of satisfaction when experiencing a strategy; if she is aware of the value a particular strategy gave her when she had one level of baseline utility, she should be able to predict what value it gives her when she starts from another level of baseline utility.

In conclusion, we have proposed a model of preference discovery that has special implications for choice under risk. Results from some existing lab experiments hint that exploration of such a model can be fruitful. However, many questions about this learning process remain unanswered, and can only be addressed with continued theory and experimental work.

References

Andersen, Steffen, Harrison, Glenn W., Lau, Morten I. and Rutstrom, E. Elisabet, 2008. "Lost in State Space: Are Preferences Stable?" *International Economic Review*, 49(3), 1091-1112.

Andreoni, James and Sprenger, Charles, 2010. "Certain and Uncertain Utility: The Allais Paradox and Five Decision Theory Phenomena." University of California San Diego, San Diego, CA.

Barron, Greg and Erev, Ido, 2003. "Small feedback-based decisions and their limited correspondence to description-based decisions." *Journal of Behavioral Decision Making*, 16(3), 215-233.

Bateman, Ian J., Burgess, Diane, Hutchinson, W. George and Matthews, David I., 2008. "Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness." *Journal of Environmental Economics and Management*, 55(2), 127-141.

Birnbaum, Michael H. and Schmidt, Ulrich, 2009, "The Impact of Experience on Violations of Independence and Coalescing." University of Kiel and California State Fullerton.

Bone, John, Hey, John and Suckling, John, 1999. "Are Groups More (or Less) Consistent Than Individuals?" *Journal of Risk and Uncertainty*, 18(1), 63-81.

Braga, Jacinto, Humphrey, Steven J. and Starmer, Chris, 2009. "Market Experience Eliminates Some Anomalies--And Creates New Ones." *European Economic Review*, 53(4), 401-416.

Braga, Jacinto and Starmer, Chris, 2005. "Preference Anomalies, Preference Elicitation and the Discovered Preference Hypothesis." *Environmental and Resource Economics*, 32(1), 55-89.

Brown, George W., 1951, Iterative solution of games by fictitious play. In: T. C. Koopmans (Ed.), *Activity analysis of production and allocation*. Wiley, New York, pp. 374-376.

Busemeyer, Jerome R. and Townsend, James T., 1993. "Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment." *Psychological Review*, 100(3), 432-459.

Bush, Robert R. and Mosteller, Frederick, 1955, *Stochastic models for learning*. Wiley, New York.

Cohen, Michele, Etner, Johanna and Jeleva, Meglena, 2008. "Dynamic Decision Making When Risk Perception Depends on Past Experience." *Theory and Decision*, 64(2-3), 173-192.

Coursey, Don L., Hovis, John L. and Schulze, William D., 1987. "The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value." *The Quarterly Journal of Economics*, 102(3), 679-690.

Cox, James C. and Grether, David M., 1996. "The Preference Reversal Phenomenon: Response Mode, Markets and Incentives." *Economic Theory*, 7(3), 381-405.

Cubitt, Robin P., Starmer, Chris and Sugden, Robert, 2001. "Discovered preferences and the experimental evidence of violations of expected utility theory." *Journal of Economic Methodology*, 8(3), 385-414.

Fischhoff, B., Slovic, P. and Lichtenstein, S., 1988. "Knowing what you want: Measuring labile values." *Decision Making: Descriptive, Normative and Prescriptive Interactions*, 398-421.

Harrison, Glenn W., Johnson, Eric, McInnes, Melayne M. and Rutström, E. Elisabet, 2005. "Temporal stability of estimates of risk aversion." *Applied Financial Economics Letters*, 1(1), 31-35.

Hertwig, Ralph, Barron, Greg, Weber, Elke U. and Erev, Ido, 2004. "Decisions From Experience and the Effect of Rare Events in Risky Choice." *Psychological Science*, 15(8), 534-539.

Hey, John D., 2001. "Does Repetition Improve Consistency?" *Experimental Economics*, 4(1), 5-54.

Horowitz, John K., 1992. "A test of intertemporal consistency." *Journal of Economic Behavior & Organization*, 17(1), 171-182.

Isen, Alice M., 2005, Positive Affect. *Handbook of Cognition and Emotion*. John Wiley & Sons, Ltd, pp. 521-539.

Jessup, Ryan K., Bishara, Anthony J. and Busemeyer, Jerome R., 2008. "Feedback Produces Divergence From Prospect Theory in Descriptive Choice." *Psychological Science (Wiley-Blackwell)*, 19(10), 1015-1022.

Kahneman, Daniel and Snell, Jackie, 1990, Predicting Utility. In: Robin M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn*. Chicago and London: University of Chicago Press, pp. 295-310.

Kahneman, Daniel, Wakker, Peter P. and Sarin, Rakesh, 1997. "Back to Bentham? Explorations of Experienced Utility." *The Quarterly Journal of Economics*, 112(2), 375-405.

Keren, Gideon and Wagenaar, Willem A., 1987. "Violation of utility theory in unique and repeated gambles." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3), 387-391.

Kingsley, David C. and Brown, Thomas C., 2010. "Preference Uncertainty, Preference Learning, and Paired Comparison Experiments." *Land Economics*, 86(3), 530-544.

Knetsch, Jack L., Tang, Fang-Fang and Thaler, Richard H., 2001. "The Endowment Effect and Repeated Market Trials: Is the Vickrey Auction Demand Revealing?" *Experimental Economics*, 4(3), 257-269.

Ladenburg, Jacob and Olsen, Soren Boye, 2008. "Gender-Specific Starting Point Bias in Choice Experiments: Evidence from an Empirical Study." *Journal of Environmental Economics and Management*, 56(3), 275-285.

List, John A., 2003. "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics*, 118(1), 41.

List, John A. and Shogren, Jason F., 1999. "Price Information and Bidding Behavior in Repeated Second-Price Auctions." *American Journal of Agricultural Economics*, 81(4), 942-949.

Loomes, Graham, Moffatt, Peter G. and Sugden, Robert, 2002. "A Microeconometric Test of Alternative Stochastic Theories of Risky Choice." *Journal of Risk and Uncertainty*, 24(2), 103-130.

Loomes, Graham and Sugden, Robert, 1998. "Testing Different Stochastic Specifications of Risky Choice." *Economica*, 65(260), 581-598.

March, James G., 1996. "Learning to be risk averse." *Psychological Review*, 103(2), 309-319.

Noussair, Charles, Robin, Stephane and Ruffieux, Bernard, 2004. "Revealing consumers' willingness-to-pay: A comparison of the BDM mechanism and the Vickrey auction." *Journal of Economic Psychology*, 25(6), 725-741.

Plott, Charles R., 1996, Rational Individual Behaviour in Markets and Social Choice Processes: The Discovered Preference Hypothesis. In: Kenneth J. Arrow and et al. (Eds.), *The rational foundations of economic behaviour: Proceedings of the IEA Conference held in*

Turin, Italy. IEA Conference Volume, no. 114. New York: St. Martin's Press; London: Macmillan Press in association with the International Economic Association, pp. 225-250.

Rozin, P., 1982, Human food selection: the interaction of biology, culture and individual experience. In: L.M. Barker (Ed.), *The psychobiology of human food selection*. AVI, Westport, Conn, pp. 225-254.

Rozin, P. and Vollmecke, T. A., 1986. "Food Likes and Dislikes." *Annual Review of Nutrition*, 6(1), 433-456.

Sarin, Rajiv and Vahid, Farshid, 1999. "Payoff Assessments without Probabilities: A Simple Dynamic Model of Choice." *Games and Economic Behavior*, 28(2), 294-309.

Shogren, Jason F., Cho, Sungwon, Koo, Cannon, List, John, Park, Changwon et al., 2001. "Auction mechanisms and the measurement of WTP and WTA." *Resource and Energy Economics*, 23(2), 97-109.

Shogren, Jason F., Shin, Seung Y., Hayes, Dermot J. and Kliebenstein, James B., 1994. "Resolving Differences in Willingness to Pay and Willingness to Accept." *The American Economic Review*, 84(1), 255-270.

Stigler, George J. and Becker, Gary S., 1977. "De Gustibus Non Est Disputandum." *American Economic Review*, 67(2), 76-90.

Thaler, Richard H., Tversky, Amos, Kahneman, Daniel and Schwartz, Alan, 1997. "The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test." *The Quarterly Journal of Economics*, 112(2), 647-661.

Thorndike, Edward L., 1898. "Animal intelligence: An experimental study of the associative processes in animals." *Psychological Monographs: General and Applied*, 2(4).

van de Kuilen, Gijs and Wakker, Peter P., 2006. "Learning in the Allais Paradox." *Journal of Risk and Uncertainty*, 33(3), 155-164.