Rational preference shifts in multi-attribute choice: What is fair?

Pradeep Shenoy (pshenoy@cs.washington.edu)
Microsoft AdCenter, Bangalore, India

Angela J. Yu (ajyu@ucsd.edu)
Department of Cognitive Science, UC San Diego, 9500 Gilman Dr. MC 0515
La Jolla, CA 92093 USA

Abstract

Humans exhibit certain systematic context-dependent preference reversals when choosing among options that vary along multiple attribute dimensions. For instance, the attraction, similarity, and compromise effects each involves a change in relative preference between two options when a third option is introduced. Previously, such effects have been attributed to irrationality or suboptimality in decision-making, or to specific architectural or dynamical constraints on cognition. We use a Bayesian model of multi-attribute choice to demonstrate that these effects naturally arise from three basic assumptions: (1) humans assess options relative to "fair market value" as inferred from prior experience and available options; (2) attributes are imperfectly substitutable, and scarce attributes are relatively more valuable; (3) uncertainty about market conditions and option values contributes to stochasticity in choice behavior. This work provides both a novel normative explanation for contextual modulation of choice behavior, and a means to predict choice as a function of past experiences and novel contexts.

Keywords: multi-attribute decision-making; preference shift; context effects; attraction effect; compromise effect; similarity effect

Introduction

Everyday decision-making often involves choosing among options that differ in multiple attribute dimensions. For example, should you buy a house that is more spacious or one that is better located? Understanding how humans make these multi-attribute decisions, and how their choices depend on the context, is an important problem in cognitive science.

Multi-attribute decision-making is particularly challenging because there is often no universal or intrinsic way to assign relative values to the different attributes. This is especially true in contexts where the decision-maker has limited experience (and thus significant uncertainty about market conditions), such as with big-ticket items like houses, or new technology like smart phones. Human choice behavior in multi-attribute problems exhibits certain systematic shifts due to context changes, such as when the relative preference between two options shift or even reverse when a third option, known as a *decoy*, is added, leading to suggestions of underlying irrationality or suboptimality (Kahneman & Tversky, 1979; Kahneman, Slovic, & Tversky, 1982; Tversky & Simonson, 1993).

In the *attraction effect* (Fig. 1A), given two similarly preferred options, A and B, the introduction of a third option Z that is similar to B, but also clearly inferior to B in one or both attribute dimensions, results in an increase in relative preference for B over A (Huber, Payne, & Puto, 1982; Heath & Chatterjee, 1995). In the *compromise effect* (Fig. 1B), when $B > A$ in one attribute and $B < A$ in another attribute, and Z has the same tradeoff but is even more extreme than B, then B becomes the "compromise" option and becomes preferred relative to A (Simonson, 1989). In the *similarity effect* (Fig. 1C), the introduction of a third option Z, that is very similar and comparable to B in both attribute dimensions, shifts the relative preference away from B to A (Tversky, 1972).

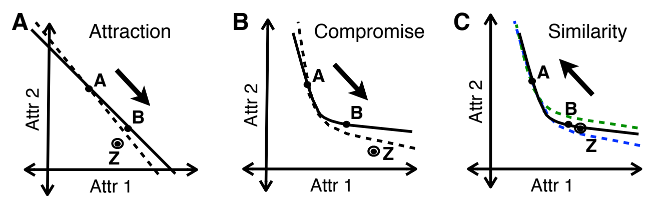

Figure 1: Three classical contextual effects in multi-attribute choice: (A) attraction effect, (B) compromise effect, (C) similarity effect. A and B are two equally preferable choices that differ in two attribute dimensions. The introduction of a third option Z induces a preference shift between A and B (indicated by arrows). Solid and dashed lines illustrate model-inferred "fair value" indifference curve before and after introducing Z.

Two broad classes of models have previously been proposed for contextual effects in multi-attribute choice behavior: (1) normative models (Marr, 1982) that are built on behavioral constraints/goals and delineated in terms of internal beliefs and assumptions (Luce, 1959; Thurstone, 1954; Luce, 1965; Tversky, 1972; Tversky & Simonson, 1993); (2) algorithmic or implementational models that explain behavioral phenomena as arising from specific architectural and dynamical constraints on neural processing (Busemeyer & Townsend, 1993; Usher & McClelland, 2004; J. S. Trueblood, 2012).

The first class of models are related to bounded rationality (Simon, 1955), but have so far been unable to explain all three contextual effects, leading to suggestions that such preference shifts reflect biases or suboptimalities in human decision-making. For example, the discovery of the similarity effect invalidated Luce's early ratio-of-strength model (Luce, 1959), and other related models that follow the simple scalability principle (Tversky, 1972). Tversky proposed the elimination-by-aspects model (Tversky, 1972) to explain the similarity effect, but it was invalidated by the discovery of the attraction effect, which violates the *regularity* principle, thus ruling out a large class of random utility models (Luce, 1965), including Thurstone's preferential choice theory (Thurstone, 1954). The compromise effect presented further complication, as no previous model could account for it, and a new context-dependent preference model (Tversky & Simonson, 1993) was only able to account for it, along with the attraction effect, by letting slip the similarity effect (Roe, Busemeyer, & Townsend, 2001).

The second class of models can account for all three effects, but are based on rather detailed and specific assumptions about neural dynamic and architecture, which have thus far not been verified experimentally, and whose computational provenance and consequences are unclear.

Here, we propose a novel rational account of multi-attribute decision-making that explains all three contextual effects. The model is grounded in three basic, empirically motivated assumptions: (1) humans make preferential choices based on relative values anchored with respect to what is perceived "fair" in the marketplace (Ariely, 2008), which is inferred from observed data, including the set of available options (Wernerfelt, 1995; Sher & McKenzie, 2011); (2) different attributes are imperfect substitutes for one another (Hicks, 1932), in particular one unit of a scarce attribute is more valuable than an abundant one; (3) uncertainty in posterior belief about "market conditions" contributes to stochasticity in preference on repeated presentations of the same options (see e.g., Debreu, 1958). We formalize these assumptions using a Bayesian generative model, and demonstrate that all three contextual effects are consequences of rational (Bayesian) inference of relative value, conditioned on the available options. In contrast to previous models, we view each decision as not only an expression of choice, but also as an opportunity for *learning* about the marketplace based on the set of options given. Thus, an individual's preference can differ in different contexts, not because of arbitrary context-dependent factors (Hsee, Zhang, Yu, & Xi, 2004; Srivastava & Schrater, 2012), but because of normative evolution of an individual's internal beliefs about the option landscape. Moreover, our model provides a means to *predict* individual and group preferences in novel contexts given past choices. In the following, we first describe the Bayesian model, followed by a comparison of simulated model behavior and empirically observed contextual effects found in the literature, and finally conclude with a discussion.

Bayesian model of relative value inference

We begin with an intuitive explanation for contextual effects before delving into the technical details of the model. While we explain the phenomena primarily in terms of consumer decision-making here, in the Discussion we will extend the model and explanation beyond choices among consumer products.

We model presented options as being drawn from a shared landscape of options, which implies that the options are *representative* of the market in some sense, and are useful for inferring general market conditions. In fact, humans often use available context to infer a reference point for valuation–for instance, in the framing effect, humans evaluate the quality of an outcome differently based on whether it is described in terms of success rates or failure rates (Sher & McKenzie, 2006). In the case of multi-attribute valuation, "fair market value" could potentially be inferred by fitting an equi-preference contour, or indifference curve (Pareto, 1927), through the presented options, where points above the line would be a "good deal", while ones below would be a "bad deal." Given the formal relationship between regression and inference (Bishop, 2006), this process is equivalent to inferring mean market value and relative attribute importance based on the samples. We first use this general intuition to explain the three effects, and subsequently present a precise generative model and inference procedure for multi-attribute choice.

In the *attraction effect*, A and B both initially lie on the inferred "fair value" indifference curve. Introducing Z, which is close to B but clearly inferior in one or both attribute dimensions, drives down the inferred "fair value" indifference line (dashed line) near B, making B appear to be a good deal (while A is still fair, and Z is worse than fair). The *compromise effect* arises from imperfect substitutability and diminishing marginal utility (Hicks, 1932) – e.g. the value of a small house in a good location would increase much more with a small increase in size than it would with a slight improvement in location. Thus, the indifference curves, including the fair value curve, should be strictly convex rather than linear. The compromise effect then naturally arises when Z is introduced, because the convex line corresponding to "fair" passes between B and Z, making B appear to be better than fair (and A fair or worse than fair). To account for the *similarity effect*, we adopt a stochastic decision policy that reflects posterior uncertainty about both market conditions and option values: the model samples from the joint posterior distribution over option values and chooses the maximally valued option. Because of the proximity of B and Z in the attribute space, inferred values of B and Z are highly correlated. For each possible setting of market conditions (family of indifference curves), B and Z tend to be both better or worse than A. This gives an overall probability of choosing A with $1/2$ probability, and choosing B (and also Z) with $1/4$ probability.

Model

$$
\begin{aligned}
&\mu \sim \Gamma(k_\mu, t_\mu)\\
&\gamma \sim \beta(a_\gamma, b_\gamma)\\
&\theta_i \sim \mathcal{N}(0, \sigma_\theta^2)\\
&v_i \sim \Gamma(k_v, \tfrac{\mu}{k_v})\\
&o_i \sim \mathcal{N}(f(v_i, \theta_i, \gamma), \sigma_o^2 \mathbf{I})\\
&f(v_i, \theta_i, \gamma) = (v_i^{\frac{1+\theta_i}{2\gamma}}, v_i^{\frac{1-\theta_i}{2(1-\gamma)}})
\end{aligned}
$$

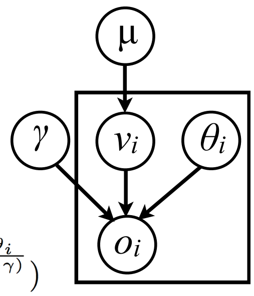

Figure 2: Bayesian generative model of relative value inference. Each two-attribute option $o_i = (x_i, y_i)$ has an underlying scalar *value* v_i, parameterized by γ, θ_i. The value v_i itself is generated from a prior distribution with mean μ, which corresponds to "fair" value.

The critical assumptions in our model are that subjects use available options to infer about the utility function and "fair market value." We assume subjects do so by inverting a hierarchical Bayesian generative model (Fig. 2), where: (1) values $\{v_i\}$ for the set of options $\{i\}$ are drawn from a prior distribution with mean μ, and (2) 2-d attribute values for each option, o_i, is generated from v_i according to a common utility function and then corrupted by observation noise. For simplicity, we use the classical Cobb-Douglas utility function (Douglas, 1976), parameterized by γ, $v_i = x_i^\gamma y_i^{(1-\gamma)}$. While more complex utility functions can be used, for example to take into account variability in the relative scaling of the two attributes, the contextual effects are not dependent on the choice of utility function, and thus not dealt with further here. To model observation noise, we first map value into an indifference curve in the attribute space by inverting the utility function, then add Gaussian noise along the indifference curve (parameterized by θ_i) and isotrophic 2-D Gaussian noise (parameterized by σ_0). We expect the main results to hold independent of the specific choices of model parameterization.

Subsequent to doing posterior inference, we assume humans choose an option by first sampling from the joint posterior $P(\mathbf{v}|\mathbf{o})$, and then (always) choosing the option with the highest sampled value. The computation of the posterior requires marginalizing over unceratinty about market conditions through a series of steps:

$$P(\mathbf{v},\mathbf{o},\mu,\gamma) \quad = \quad p(\mu)p(\gamma)\Pi_i[\int_{\theta_i} p(\theta_i)P(v_i|\mu)P(o_i|\theta_i,v_i,\gamma)]$$

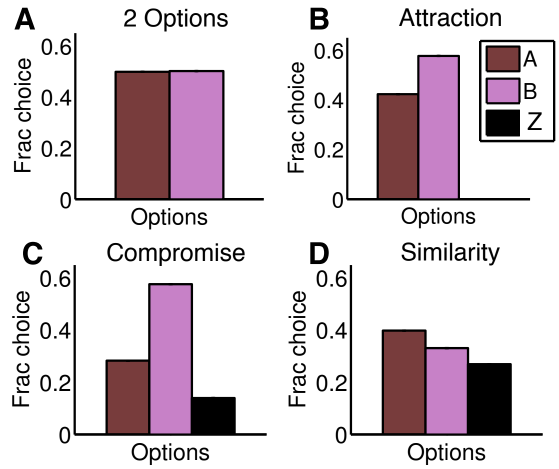

Figure 3: Preference shifts as rational inference. (A) Model chooses A and B equally when there are only two options. (B) Attraction: introducing the inferior Z makes B more preferable to A (Z is almost never chosen). (C) Compromise: introducing an extreme option Z makes B more preferable to A. (B) Similarity: introducing Z, highly similar to B, makes B less preferable to A.

$$P(\mathbf{v}|\mathbf{o}) \quad \propto \quad \int_{\mu,\gamma} P(\mathbf{v},\mathbf{o},\mu,\gamma)$$

Simulation details

The parameter settings for our simulations were as follows: $(k_\mu, t_\mu) = (1, 100); (a_\gamma, b_\gamma) = (2, 2); \sigma_\theta = 20; k_v = 20; \sigma_o = 2$ (see Fig. 2). The Gamma distributions were parametrized using parameters for *shape* (k_μ), and *scale* (t_μ), and the mean of the corresponding distribution is given by their product (e.g., $k_\mu \cdot t_\mu$). Accordingly, the mean of the prior distribution over μ is 100, and the shape parameter encodes a broad uncertainty about the true value of μ (see Fig. 4).

We finely discretized each of the variables in our model to calculate the relevant posterior distributions numerically (analytical solutions do not exist). The option values used (see Fig. 1) were as follows: $A = (40, 60), B = (60, 40)$; attraction: $Z = (30, 50)$, compromise: $Z = (80, 20)$, similarity: $Z = (65, 35)$.

Results

Preference shift as option-based value inference

As Fig. 3 shows, simulations of our model reproduces all three contextual effects: attraction, compromise, and similarity. In particular, the model reproduces violation of regularity in attraction effect that is also seen in human data. In all three cases, although options A and B are equally preferred when presented as a pair (Fig. 3A), the presence of a third (decoy) option Z changes this relative preference (Fig. 3B-D). This shift in preference depends on the relationship between the precise attribute values of the decoy relative to those of the two original two options (see Fig. 1). All three contextual effects were obtained using the same model setting, except for the position of the decoy Z. Thus, these contextual shifts in preference can indeed be direct consequences of normative inference about relative values, conditioned on both prior beliefs and the available options. Note that the main results hold over a wide range of parameter settings and are not sensitive to the particular parameterization of the model.

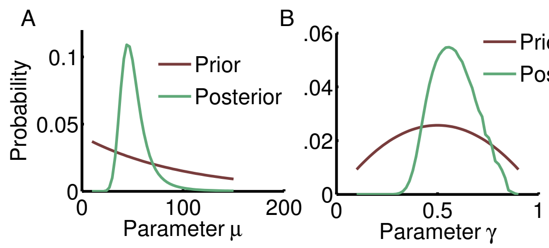

Figure 4: Posterior distributions over model variables for the *compromise effect*. (A) Marginal posterior distribution over what constitutes "fair" in the market, parameterized by μ, before (red) and after (green) introducing Z. (B) (A, B) Marginal posterior distribution over the shape of the family of indifference curves, parameterized by γ, before (red) and after (green) introducing Z.

We explore the *compromise effect* in more detail to illustrate the inner workings of our model. The joint inference over (μ, γ) is reflected in the shape of the equi-preference contours and the probability of each contour being "fair" (Fig. 5): colored bands represent indifference curves for the MAP estimate of γ and a range of values of μ, and the color indicates the probability of that band representing fair market value. When only options (A, B) are presented (Fig. 5A), the fair market value contour passes through both A and B; when Z is introduced (Fig. 5B), the contours shift so as to make B better than fair (and A fair).

Next, we examine the properties of the inferred joint posterior distribution $P(\mathbf{v}|\mathbf{o})$, illustrated in Fig. 6. Shown in panel A are the marginal value distributions for the 3 options A, B, Z in the compromise effect. Consistent with Fig. 5, the inferred value distributions show a clear ordering, with option B having the highest expected value (Fig. 6B). However, the marginal distributions (panel A), and expected values (panel B) do not capture implicit correlations among inferred values induced by the shared (marginalized) variables μ and γ. In fact, our model samples from the joint posterior value distribution and selects the highest value in each sample. Fig. 6C shows the the empirical probability of each

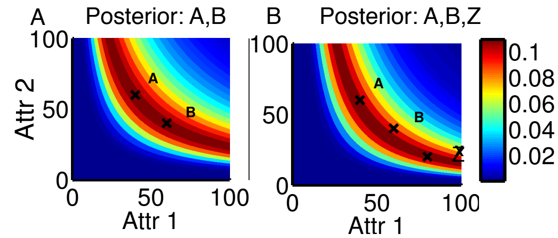

Figure 5: Joint marketplace and value inference in the compromise effect: (A) given only A and B as options, (B) given A, B, Z. Each colored band represents an equi-preference contour (indifference curve) corresponding to the MAP estimate of γ, with its color indicating the probability of its being the mean market value.

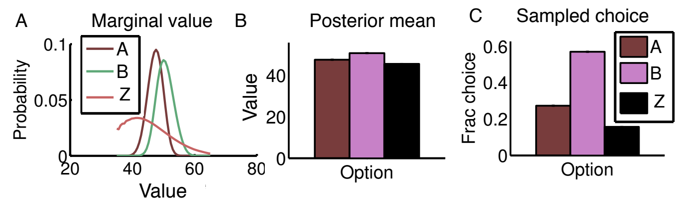

Figure 6: Value inference and sampling in *compromise effect*. (A) Posterior distributions over option values. (B) Mean posterior value. (C) Empirical choice distribution based on samples (n=1000) from the joint posterior distribution over option values.

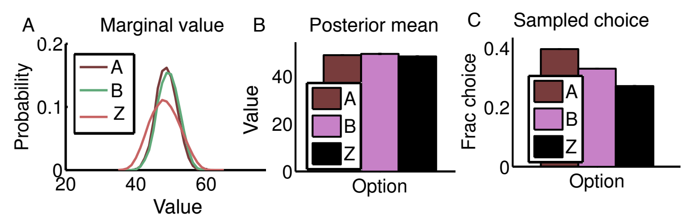

Figure 7: Value inference and sampling in *similarity effect*. See Fig. 6.

option being chosen. In the joint distribution, A and Z are positively correlated in inferred value, and, as a result, our model strongly prefers the compromise option B over A, stronger than would be suggested by the marginal distributions alone.

Correlations in the joint posterior value distribution is particularly important also for generating the similarity effect (Fig. 7), where the marginal distributions and mean values for A and B are indistinguishable from each other, but the sampled preference for A is much higher, due to a strong positive correlation between the inferred values of B and Z.

Model predictions

Our model makes a number of experimentally testable predictions about multi-attribute choice behavior. Since presented options not only influence the immediate choice but also general beliefs about general market conditions, our model predicts systematic consequences in future choice behavior based on experienced choice history. For instance, subjects exposed to a number of choices between options generally higher in one attribute may correspondingly learn a γ that discounts this attribute more – resulting in a smaller attraction effect for a decoy that is inferior to B in this attribute dimension compared to the other. There is some evidence that subjects show such "context-dependent utility functions" (Drolet, Simonson, & Tversky, 2000).

Another arena for experimental exploration suggested by this work is the transition among the different effects due to the precise positioning of the options in the attribute space: for instance, the "similarity" decoy in Fig. 1C could well turn into a "compromise" decoy in Fig. 1B, if it were far enough from B. Thus, one prediction of our model is that as the decoy Z is moved away from the option B, while maintaining a rough tradeoff between the two attributes, the contextual effect changes from similarity effect to compromise effect. That is, if the decoy were exactly the same as B, preference should shift *away from* B, but as the decoy is moved further apart, preference should shift *toward* B. In an analogous manner, we expect to see a smooth transition between the similarity and attraction effects as the decoy is moved away in the orthogonal, dominated direction. Fig. 8 shows that model simulations conform to these expectations: as the decoy is moved further along non-dominated (panel A) or dominated (panel B) directions, the model predicts a gradual evolutation from a similarity effect to the compromise and attraction effects, respectively.

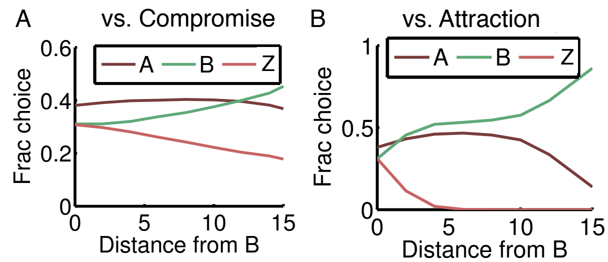

Figure 8: Transitions in relative preference. The three effects are related to each other by the magnitude of the distance between options. (A) When a third option Z, initially identical to B, is moved away in a non-dominated direction, relative preference changes from favoring A (similarity effect) to favoring B (compromise effect). (B) When Z is moved away from B in an orthogonal, dominated direction, preference changes from favoring A (similarity effect) to favoring B (attraction effect).

Discussion

We presented a normative Bayesian model for why human subjects exhibit apparently irrational choice behavior in multi-attribute decision-making. We showed that violations of the simple scalability and regularity principles need not be reflections of an irrational or sub-optimal decision or valuation process, but rather rational consequences of a decision-maker who is trying to optimize choice in a relativistic system anchored to what is *perceived* to be fair. We used a normative, hierarchical Bayesian generative model to demonstrate how the set of options themselves can be used to infer about the landscape of available options, such as how value is distributed in the market, how the multi-dimensional observed attribute space is mapped to the scalar value representation, and the distribution of observation noise. Although the language of this paper primarily focuses on consumer decision-making, the model can be extended to a much broader range of multi-attribute choice behavior, whenever the observer has uncertainty about how to combine two attributes in order to compare the options. In future work, we plan to extend the current model to explore some non-consumer choice tasks known to exhibit context effects (Choplin & Hummel, 2005; J. S. Trueblood, 2012; J. Trueblood, Brown, Heathcote, & Busemeyer, n.d.).

Our approach contrasts with the class of models that explain contextual effects based on specific architectural or dynamic constraints on neural processing. One example is the decision field theory (DFT) model (Busemeyer & Townsend, 1993), which assumes that the strength of preference for each option is driven by a noisy, accumulative input and dynamical switching of "attention" among different attribute dimensions, as well as "lateral inhibition" between the different units. A related model (J. S. Trueblood, 2012), an extension of the multi-attribute linear ballistic accumulator model (Brown & Heathcote, 2008), employs attentional switching, a contrast mechanism (related to lateral inhibition), and sensitivity to indifference/dominance. A third model, the competing accumulator model (Usher & McClelland, 2004), assumes loss aversion in addition to attentional switching and lateral inhibition. The various overlapping and nonidentical assumptions of these process models are difficult to verify experimentally, and their computational provenance/constraints are not well understood. This is not to say that such mechanistic models are not useful. Ultimately, to understand how the brain implements multi-attribute choice, we need multiple levels of analysis (Marr, 1982) that integrate both normative and mechanistic explanations. In this vein, our work complements existing work by helping to frame and constrain mechanistic models.

Although the model presented here succinctly and rationally accounts for contextual effects in multi-attribute choice behavior, it is clearly not a complete theory of human preference choice. In particular, the simple model presented here has no means of accounting for individual differences according to taste. A natural way this arises is when people bring in different previous experiences and thus prior beliefs about the market. However, this cannot be the whole story, as any prior difference would be overwhelmed by sufficient data, and yet people who have repeated exposure to the same choices do not always converge in their preferences (e.g. office workers who eat out at the same set of neighborhood restaurants day after day). An important line of future enquiry is how individual differences in preference may arise and persist in the face of mounting, common experiences.

References

Ariely, D. (2008). *Predictably irrational*. Harper Collins.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Brown, S., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive psychology*, *57*(3), 153–178.

Busemeyer, J., & Townsend, J. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review*, *100*(3), 432–459.

Choplin, J., & Hummel, J. (2005). Comparison-induced decoy effects. *Memory & Cognition*, *33*(2), 332–343.

Debreu, G. (1958). Stochastic choice and cardinal utility. *Econometrica*, *26*, 440–444.

Douglas, P. (1976). The Cobb-Douglas production function once again: Its history, its testing, and some new empirical values. *The Journal of Political Economy*, *84*, 903–915.

Drolet, A., Simonson, I., & Tversky, A. (2000). Indifference curves that travel with the choice set. *Marketing Letters*, *11:3*, 199–209.

Heath, T., & Chatterjee, S. (1995). Asymmetric decoy effects on lower-quality versus higher-quality brands: Meta-analytic and experimental evidence. *Journal of Consumer Research*, *22*, 268–284.

Hicks, J. R. (1932). *The Theory of Wages*. London, UK: Macmillan.

Hsee, C., Zhang, J., Yu, F., & Xi, Y. (2004). Lay rationalism and inconsistency between predicted experience and decision. *Journal of Behavioral Decision Making*, *16*.

Huber, J., Payne, J., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*(1), 90–98.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. New York, NY: Cambridge Univ Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, *47*(2), 263–292.

Luce, R. (1959). *Individual choice behavior: a theoretical analysis*. New York: Wiley.

Luce, R. (1965). Preference, utility, and subjective probability. In R. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 3, pp. 249–410). New York: Wiley.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

Pareto, V. (1927). *Manual of political economy Tr. by Ann S. Schwier*. London: Macmillan.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological Review*, *108*(2), 370–392.

Sher, S., & McKenzie, C. (2006). Information leakage from logically equivalent frames. *Cognition*, *101*(3), 467–494.

Sher, S., & McKenzie, C. R. (2011). Levels of Information: A Framing Hierarchy. In *Perspectives on framing*. Psychology Press.

Simon, H. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, *69*(1), 99–118.

Simonson, I. (1989). Choice based on reasons: the case of attraction and compromise effects. *Journal of Consumer Research*, *16*(158-74).

Srivastava, N., & Schrater, P. (2012). Rational inference of relative preferences. *Advance in Neural Information Processing Systems*, *25*.

Thurstone, L. (1954). The measurement of values. *Psychological Review*, *61*, 47–58.

Trueblood, J., Brown, S., Heathcote, A., & Busemeyer, J. R. (n.d.). Not just for consumers: Context effects are fundamental to decision-making. *Psychological Science*.

Trueblood, J. S. (2012, October). Multialternative context effects obtained using an inference task. *Psychonomic Bulletin & Review*, *19*(5), 962–8.

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

Tversky, A., & Simonson, I. (1993). Context dependent preferences. *Management Science*, *39*, 1179–89.

Usher, M., & McClelland, J. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological review*, *111*(3), 757–760.

Wernerfelt, B. (1995). A rational reconstruction of the compromise effect: Using market data to infer utilities. *The Journal of Consumer Research*, *4*(2), 627–633.