Decision and Choice: Luce's Choice Axiom

Timothy J. Pleskac*

Department of Psychology, Michigan State University

September 23, 2012

Abstract

Luce's choice axiom (LCA) is a theory of individual choice behavior that has proven to be a powerful tool in the behavioral sciences for over 50 years. LCA is grounded in two fundamental properties: choice is probabilistic and the probability of choosing an option from one set of alternatives is related to the probability of choosing the same option from a different set. This entry reviews the basic properties of LCA. It also discusses crucial tests of predictions derived from LCA regarding the role of context. The historical connections of LCA and crucial applications of LCA are also reviewed.

Keywords: decision; choice; psychology; economics; preference; learning; axiomatic; Luce; probability; stochastic; context; mathematical model;

1 Introduction

A subject in a psychology experiment (animal or human) identifies which of two stimuli are brighter; a diner at a restaurant selects a meal from a menu; an eyewitness determines a face in a line up. These are all specific examples

*Email: pleskact@msu.edu; Phone: (517) 303-5743. A grant from the National Science Foundation (0955410) supported this work. The author wishes to thank Mitchell Uitvlugt and Patrycja Zdziarska for help in editing the submission.

of a general behavior called choice where an individual selects one option from a larger set of alternatives.

Owing to the generality of choice behavior, many different areas of the social sciences have sought a mathematical description of choice. Luce's (1959) choice axiom (LCA) is one such description of individual choice behavior. As Estes (1997) describes it, Luce, in developing a theory of individual choice behavior, did not take the standard approach of psychology.[1] Instead of working from the bottom-up accumulating empirical data on a phenomenon of interest and eventually arriving at a general theory of choice behavior, Luce took the approach of theoretical physics and worked from the top-down. He employed intuition and reason to find a set of general assumptions that if true would allow the development of a theory and mathematical model useful for interpreting and understanding choice behavior. In short, LCA helped usher in the axiomatic method into psychology and more broadly the social sciences.

Luce started with two basic tenets of individual choice behavior: (a) it is probabilistic; and (b) the probability of choosing an option from one set of alternatives is related to the probability of choosing the same option from a larger set of alternatives. The tenet that choice behavior is probabilistic is in contrast to assuming choice is deterministic. To understand this distinction, consider a situation where an agent chooses an option x from the set of alternatives T (e.g., possible entrees on a menu). Focusing on the special case of 2-alternative forced choice or paired comparison, deterministic theories assume a binary preference relation $\prec$ that is either true or false for any pair of alternatives, either $x \prec y$, $y \prec x$, or $x \sim y$ for all $x, y \in T$ where $\sim$ is indifference.

A probabilistic choice theory, like LCA, makes the more psychologically plausible assumption that it is only with some probability that an alternative is selected. Focusing again on a paired comparison, a probabilistic choice theory specifies a probability function $P(x, y)$ that maps each pair of alternatives into the closed interval $[0, 1]$. LCA is a multi-alternative choice theory so it is not limited to paired comparisons, but is applicable in the more general choice situation specifying the probability that x will be chosen from the set T, $P_T(x)$. Regardless, as a probabilistic theory, LCA adheres

[1]Evidentally, a draft of the book was distributed as a pamphlet with a red cover during the Social Science Research Council Summer Instiute on Mathematical Training in the Social Scielces at Stanford in the summer of 1957. The pamphlet and the ideas in it were subsequently referred to as the "red menace" (Estes, 1997).

to the axioms of probability theory. These axioms are

1. For $S \subset T, 0 \leq P_T(S) \leq 1$
2. $P_T(T) = 1$
3. If $R, S \subset T$ and $R \cap S = 0$, then $P_T(R \cup S) = P_T(R) + P_T(S)$.

The probability axioms constrain each of the measures P_T. The problem is that the axioms themselves offer no connection between measures over different sets composed of some of the same alternatives; yet, such a connection seems necessary for a theory of choice. Anecdotally, the probability that one chooses pan seared walleye from the dinner menu at a restaurant certainly seems related to the probability that he or she will choose the same entree from the lunch menu. This observation motivates a second tenet of choice behavior: how decision makers select an option from a smaller set of alternatives is related to how they would choose when the same option is in the context of a larger set of alternatives and vise versa. LCA formalizes this assumption.

2 The Choice Axiom

The axiom has two parts.

Part 1: If $P(X, Y) \neq 0,1$ for all x, $y \in T$, then for $R \subset S \subset T$

$$P_T(R) = P_S(R) P_T(S) \tag{1}$$

Part 2: If $P(X, Y) = 0$ for some x, $y \in T$, then for $S \subset T$

$$P_T(S) = P_{T-\{x\}}(S - \{x\}) \tag{2}$$

Working backwards, Part 2 is more or less a housekeeping assumption. It allows alternatives that are never chosen in pairwise choices to be deleted from S without impacting the choice probabilities. If salted whitefish is never chosen in pairwise choices with trout, then in choices between salted whitefish, trout, and walleye, salted whitefish can be safely deleted reducing the choice to trout or walleye.

Part 1 states that the probability of selecting the set of alternatives R (e.g., walleye or trout) from T is equal to the probability of selecting R

from S (e.g., walleye, trout, or salmon) multiplied by the probability of selecting S from T. When $S = \{x, y\}$ then the probability measure P_S reduces to a pairwise choice probability mentioned earlier. In other words, Part 1 addresses the issue of formally relating the choice probabilities for subsets of T to those of T itself. Another way to see this is to rewrite Equation 1 as a conditional probability

$$P_T(R|S) = P_S(R) = \frac{P_T(R)}{P_T(S)} \tag{3}$$

where $R \subset S \subset T$ and $P_T(S) > 0$. Thus, the probability of selecting R from S equals the conditional probability that R is chosen from S when the full set T is available.

To be clear, LCA is not an additional probability axiom. Rather, it is an empirical hypothesis that has a number of strong consequences. These consequences have been tested and in some places have been shown not to hold under specific circumstances. Nevertheless, LCA and the mathematical model it implies have proven quite useful in modeling choices whether they be higher level economic decisions or lower level perceptual choices. It has also motivated similar approaches in modeling probability judgments (Tversky & Koehler, 1994).

3 Implications

Most well known is that LCA implies that there is a numerical ratio scale v over T such that for every $x, y \in S \subset T$ the probability of choosing x can be found with the following mathematical model

$$P_S(x) = \frac{v(x)}{\sum_{y \in S} v(y)}. \tag{4}$$

Psychologically $v(x)$ is a response strength associated with the alternative x. One might see from Equation 4 that one measure of response strength is the observed choice probability $v(x) = P_T(x)$. Other scales can be created by multiplying $P_T(x)$ by a positive constant k , $v(x) = k \cdot P_T(x)$. This means that v is a ratio scale of response strength, a provocative idea in psychology that is beset with ordinal scales.

Equation 4 is an intuitive way to model choice probabilities. So much so that others have proposed it as a model of choice. For example, Bradley

& Terry (1952) proposed a model consistent with Equation 4 for pairwise choices (see also Clark, 1957). This straightforward way to model choices via response strength also exposes why LCA has proven important in modeling choices in a variety of domains.

Using Equation 4, one can also see that

$$\frac{P(x,y)}{P(y,x)} = \frac{P_S(x)}{P_S(y)}. \tag{5}$$

This result is because both sides of Equation 5 are equal to $\frac{v(x)}{v(y)}$. This relationship is sometimes called the *constant ratio rule*. The constant ratio rule is a probabilistic form of what is called independence from irrelevant alternatives (Arrow, 1951). It implies that when LCA holds for a set of alternatives T and its subsets the ratio $P_S(x)/P_S(y)$ is independent of S . In other words, the ratio of the probability of choosing one alternative to the probability of choosing another should be constant regardless of the set of options (or context).

This constant ratio rule is revealing in terms of the strengths and weakness of LCA. A key strength is that pairwise choice probabilities can be predicted from choice probabilities observed in a larger set. A potential weakness is that psychologically the context of the situation (e.g., similarity to other choice alternatives) should impact the probability of choosing an option in some manner. Such a weakness speaks to the validity of LCA, and we will return to this idea in the validity section.

LCA also constrains the possible pairwise probabilities. This constraint is known as the product rule which states that for alternatives x, y, z

$$P(x,y) \cdot P(y,z) \cdot P(z,x) = P(x,z) \cdot P(z,y) \cdot P(y,x). \tag{6}$$

To see how the product rule works, note that the left hand and right hand sides contain complementary pairwise choice probabilities. Consequently, dividing the left hand side by the right hand side produces

$$\frac{P(x,y)}{P(y,x)} \cdot \frac{P(y,z)}{P(z,y)} \cdot \frac{P(z,x)}{P(x,z)} = 1. \tag{7}$$

Each of the ratios of choice probabilities can, in turn, be rewritten as a ratio of response strengths by Equation 4. This allows each response strength to be cancelled off resulting in the product of the ratios equalling 1. The product rule is another independence condition that implies the response strength for an alternative is the same regardless of the other alternative it is paired with. Just as with the constant ratio rule, the product rule identifies the strengths and potential weaknesses of LCA. The strength is that the product rule allows one to predict the pairwise choice probability $P(x, z)$ if $P(x, y)$ and $P(y, z)$ are both known. The weakness is that again psychologically it seems that the context matters. Again the similarity between two alternatives seems like it should impact the response strength of an option.

4 Historical Connections

LCA has a number of connections to other choice models. Perhaps most prominent is the connection to L.L. Thurstone's (1927) law of comparative judgment. This theory assumes that alternatives have numerical psychological values that vary randomly from one choice occasion to another (i.e., the values are random variables). The decision maker is assumed to compare these values in order to make a choice. More precisely, if the psychological value of the alternatives $x_1, ..., x_N$ are the random variables $U_1, ...U_N$, the probability of choosing x_i from any set S is the probability that U_i is currently the largest of the random variables $\{U_j : x_j \in S\}$. The most prominent version of Thurstone's theory is Case V, which assumes the psychological values are normally, independent and identically distributed. This means U_i can be rewritten as $u_i + \epsilon_i$, where u_i is a real constant (i.e., a scale value) and ϵ_i is normally distributed with a mean of 0 and variance of 1, $\epsilon_i \sim N(0, 1)$.

Luce (1959) showed that LCA and Case V make nearly identical predictions for pairwise choice probabilities: the numerical predictions were off by at most .02. In fact, if the differences between the psychological values have a logistic distribution rather than a normal distribution, then Thurstone's law of comparative judgment and LCA make identical predictions (Adams & Messick, 1957). What Holman and Marley (cited in Luce & Suppes, 1965) subsequently showed is that a logistic distribution of differences arises in Thurstone's (1927) framework if the normal random variables ϵ_i are replaced by independent random variables that have a double exponential probability distribution.

Yellott (1977), in the spirit of Luce's top-down axiomatic approach, identified an intuitively plausible assumption that would lead one to conclude the random variables ϵ_i are double exponentially distributed. The assumption is grounded in the property that that the asymptotic distribution of the maximum of n independent, identically distributed random variables is the double exponential when the underlying distribution has an upper exponential tail. Using this property, Yellott proposed a thought experiment in which there are three basic alternatives, for example, beverages on a table: coffee, tea, and milk. Instead of 1 cup of each on the table, consider a situation like a party where there are n identical cups of coffee, n cups of tea, and n glasses of milk. One would expect that the probability of choosing any one of the beverages should be independent of n or how many replicas there are. In fact, if this invariance assumption is met, then the distribution in a Thurstone model must be a double exponential and, as a result, LCA must hold. In the words of Luce (1977), "Yellott's condition is so compelling that this theorem means Thurstone's model....and the choice axiom stand or fall together" (p. 218).

LCA also connects with a number of other models in experimental psychology. This came primarily via what is called the similarity choice model (Estes, 1997). This model is a generalization of the choice model in Equation 4 that Luce (1959) developed to account for response bias and later fully developed in Luce (1963) (see also Shipley, 1960). The similarity choice model decomposes the response strength v into two ratio scales representing two constructs of high interest to psychologists: discriminability and response bias. To see the extension, consider the situation where a participant in a psychology experiment must choose a response R_j upon presentation of a stimulus V_i (e.g., identify a letter upon the presentation of an obscured letter stimulus). The similarity choice model states the response probability is

$$P_{i,j} = \frac{M_{i,j} \cdot b_j}{\sum_k M_{i,k} \cdot b_k}. \tag{8}$$

The scale value $M_{i,j}$ represents the similarity between stimulus V_i and the stimulus corresponding with the response j , V_j . The other scale value b_j measures a subject's bias to choose response j regardless of the stimulus.

It should be noted that the similarity choice model is a generalization of the choice model (Equation 4), but not necessarily the axiom itself (Luce, 1963). Despite this restriction the similarity choice model can easily have some of the same context-free assumptions of LCA. The similarity choice model provided researchers with an alternative and more convenient means to model discriminability and response bias than with signal detection theory (Tanner & Swets, 1954). This convenience facilitated the incorporation of the similarity choice model as a choice rule into more complex models of cognitive processes including identification, recognition memory, categorization, and categorization learning (see Estes, 1997).

5 Validity of LCA

The value of the axiomatic approach that underlies LCA is that it establishes precise and testable conditions that help establish when a theory is applicable and when it is not. In the case of LCA, examples were almost immediately identified that questioned its validity.

One of the first examples came from Gerald Debreu (1960) in his published review of *Individual Choice Behavior*.[2] The example is as follows. Consider a classical music lover who is choosing between a recording of the Debussy quartet (option D) and two different recordings of Beethoven's Eighth Symphony (option B_1 and option B_2). For this example, assume the two Beethoven recordings are of equal quality (e.g., played by the same orchestra, but under the direction of different conductors). Lets assume the music lover is torn between adding a Debussy or a Beethoven recording to his collection. Therefore, in pairwise choices $P(B_1, B_2) = P(D, B_1) = P(D, B_2) = 1/2$. LCA thus implies that when presented with the set S of all three options $\{D, B_1, B_2\}$ simultaneously that $P_S(D) = 1/3$. However, this conclusion is not quite right intuitively. When the choice is between all three it seems that the choice is between Debussy or Beethoven. Put differently, $P_S(D)$ should remain closer to 1/2 while choice probability for the two Beethoven recordings should be split equally moving them closer to 1/4. Empirically, at least in preferential tasks, this is what happens and Tversky (1972) called this the similarity effect.[3]

The similarity effect is inconsistent with the constant ratio rule of LCA (Equation 4). The implication of the similarity effect is actually stronger in that it is inconsistent with any probabilistic choice theory that assumes choice probability is a monotone function of scale values also known as simple scalability models (Tversky, 1972).[4]

Generalizations of LCA can be developed to account for the similarity

[2] A famous adaptation of Debreu's example is the red-bus–blue-bus problem in transportation research (Train, 2003).

[3] Tversky (1972) found violations when subjets were asked to make choices between college applicants and gambles. However, the violations in a perceptual task of judging

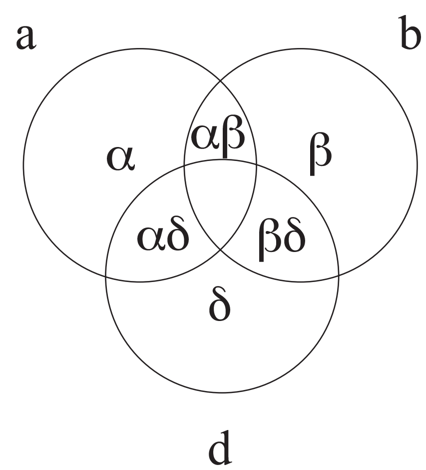

Figure 1: A graphical representation of 3 alternatives each composed of a set of unique and shared aspects.

effect. For example, Tversky (1972) proposed an alternative choice process called elimination by aspects where alternatives are sets of attributes or aspects. Figure 5 depicts a three alternative set $T = \{a, b, d\}$. The aspects $\alpha\, \beta$ and δ are unique to alternatives a, b, and d , respectively. The other aspects $\alpha\beta, \alpha\delta$, and $\beta\delta$ are shared. Each aspect has a scale value u representing its utility. Because of the elimination by aspect assumption of the theory, aspects shared among all alternatives are not used in this choice rule, hence the central area in Figure 5 is unlabeled.

The basic idea in the theory is that an aspect is attended to with some probability (e.g., Beethoven). This probability is proportional to its utility relative to the sum of all the utilities K. For example, the probability of selecting $\alpha\beta$ is $u(\alpha\beta)/K$. Alternatives that do not have that aspect (e.g., d) are then eliminated and the probability of selecting say a is the pairwise choice probability scaled from the remaining aspects

$$P(a,b) = \frac{u(\alpha) + u(\alpha\delta)}{u(\alpha) + u(\alpha\delta) + u(\beta) + u(\beta\delta)}. \tag{9}$$

In other words, elimination by aspects is consistent with one's intuitive reac-

---

dot numeroscity were not statistically significant.

[4]Similarity between alternatives also impacts pairwise choice probabilities and leads to violations of the product rule and other independence assumptions (Rumelhart & Greeno, 1971; Tversky & Russo, 1969).

tion to the Debreu's (1960) Debussy/Beethoven example: after an alternative is eliminated choice probability gets divided among the remaining alternative.

If an aspect is attended to that is unique to a particular alternative (e.g., α), then the corresponding alternative (e.g., a) is selected with probability 1. Thus, the probability of selecting alternative a from T is

$$P_T(a) = \frac{u(\alpha) + u(\alpha\beta) \cdot P(a,b) + u(\alpha\delta) \cdot P(a,d)}{K}. \tag{10}$$

Elimination by aspects is one hypothesis of how the similarity between the alternatives (i.e., shared aspects) interacts with the choice process. Using Figure 5, one can also see how elimination by aspects is a generalization of LCA: if alternatives are pairwise disjoint (i.e., no shared aspects between pairs of alternatives) then the predicted choice probabilities are consistent with LCA (Tversky, 1972).

More broadly, the similarity effect demonstrates that context matters: the response strength of an alternative depends on the options that are in the choice set. Over the years, other context effects have been identified and studied that further attest to the importance of context (Rieskamp et al., 2006). A second context effect that questions the validity of LCA (as well as elimination by aspects) is known as the attraction effect. The attraction effect can be illustrated with an example from Simonson & Tversky (1992). They gave one group of subjects a choice between \$6 and a nice Cross pen. The pen was chosen by 36% of the subjects while the remaining 64% chose \$6. A second group was given a choice between three options: \$6, a nice Cross pen, or another less attractive pen. As one might expect, only 2% chose the less attractive pen. Yet, the mere addition of this asymmetrically dominated alternative boosted the proportion of subjects who chose the Cross pen to 46% (an increase of 33%). This behavior -no doubt well known among marketers and salespeople- does not seem quite right from the view point of LCA and, for that matter, most probabilistic theories of choice: adding an alternative to a choice set of even minuscule value should reduce choice probabilities for the all the alternatives, not increase them. This condition is known as regularity. Stated formally for $x \in S \subset T$, $P_S(x) > P_T(x)$. The attraction effect demonstrates that in particular situations individuals violate regularity.

A third context effect is the compromise effect. This effect can be demonstrated with another example from Simonson & Tversky (1992) where participants made choices between hypothetical cameras. Camera A was high in quality and price while Camera B was low in quality and price. Cameral C had intermediate quality and price. In pairwise probabilities, $P(A|\{A,C\} = P(C|\{A,C\})$. However, when Camera B at the other extreme is added to the choice set $P(C|\{A,C,B\}) > P(A|\{A,C,B\})$. The compromise effect is another violation of independence from irrelevant alternatives. It questions the validity of LCA as well as elimination by aspects (for a proof see Rieskamp et al., 2006).

These empirical effects identify limitations of LCA as a model of human choice behavior. In particular, they demonstrate a third tenet of choice behavior is that context does matter and it shape how individuals value an alternative. There are different psychological explanations for the role of context. Nevertheless, LCA and the choice rule that it implies have persisted as a useful model of individual choice behavior. There are at least two reasons for this. First, the basic idea that choice probability is monotonic with a response strength is intuitively quite plausible and one that has generally persisted even as alternative models have been proposed to account for context effects. Second, the axiom and the choice rule are elegantly simple, allowing for both easy implication and easy explanation.

6 Applications in Preferential Decisions

As the previous section examining the validity of LCA illustrates, there has been great interest in applying LCA to preferential choices. These are choices made over money and other alternatives (e.g., cameras, records, commuting routes, etc.) where there is no objectively correct answer. Luce (1959) very much anticipated applications of LCA and the choice rule in this domain, but great strides were made when (c.f., McFadden, 1976, 2001) developed an economic version of Luce's choice model in the *multinomial logit* model. Formally, the probability of choosing alternative x from set S is

$$P_S(x) = \frac{exp[u(x)]}{\sum_{y \in S} exp[u(y)]}, \tag{11}$$

where $u(x)$ is the utility of alternative x and is a linear function of the attributes of x .

The multinomial logit model proved useful in economic data analysis, but it is also connected to what are called random utility models. This connection perfectly echoes the connection between LCA and Thurstonian choice models discussed earlier. Random utility models -akin to Thurstonian models- assume the utility of an alternative is a random variable. McFadden (1976), independent from Yellott (1977), showed that if the linear combination of attributes has an error component that is identically and independently distributed following an extreme value distribution (i.e., double exponential), then the result is the multinomial logit model. The multinomial logit model has become one of the most common random utility models (Train, 2003).

LCA has also contributed heavily in the further development of descriptive theories of preferential choice (Rieskamp et al., 2006). In this area, the formal attributes of LCA have proven indispensable in forming testable hypotheses about how individuals make a choice. Doing so has validated the core tenets of LCA (i.e., choice is probabilistic and there is a connection between the propensity to choose alternatives between contexts). At the same time, as discussed in the previous section, LCA very much helped uncover other tenets such as the context dependence of choice. Psychologists and decision scientists have, in turn, sought to develop other formal theories that better account for these tenets. Examples (among others) include the context-dependent advantage model (Tversky & Simonson, 1993), decision field theory (Roe et al., 2001), and the leaky accumulator model (Usher & McClelland, 2004). The latter two theories perhaps represent the greatest departure from LCA in that they are grounded in a sequential sampling account of choice. Such a move allows for the theories to describe both choice probabilities and the time taken to make a choice. This is an important step as perhaps a fourth tenet of choice behavior is that the process of making a choice takes time, a property LCA is silent on.

One final application of LCA in preferential choices comes in the area of strategic games and game theory (McKelvey & Palphrey, 1995). Strategic games are situations when entities must decide how to act during interactions with other entities. Examples include two tennis players opposing each other in a match, roommates waiting for the other to purchase kitchen supplies, or research and development expenditures by competing pharmaceutical companies. These are all examples of interactions where how one entity decides to act affects the outcomes of the other and vise versa. This interaction is what distinguishes strategic games from the individual decisions where the outcomes an individual receives do not depend on the decisions of others. In game theory, each individual chooses among an action or strategy based on the utility of the outcomes that are expected to come from employing that strategy. Typically the choice between strategies in game theory is deterministic. McKelvey & Palphrey (1995) adapted the multinomial logit model (Equation 11) to move away from the deterministic assumption to one where players probabilistically choose among their strategies and assume their opponents do so as well.

The advantage of adapting a probabilistic choice rule in strategic games is two-fold. The first advantage is that the probabilistic choice rule provides a more psychologically plausible assumption than a deterministic selection of a strategy. This allows the best equilibrium strategy to be played more often than others (but not always). The second and related advantage is that this can aid in data analysis of games where observed behavior is probabilistic allowing measurement of interesting phenomena like learning during repeated play.

7 Applications in Learning

Luce (1959) also discussed LCA's application in modeling learning. Learning has long been of interest in the area of choice whether it be modeling the behavior of rats in a T-Maze as they learn a route to food or more recently a participant in a psychology experiment choosing repeatedly between gambles that deliver a reward and/or punishment with unknown probabilities. Luce's idea called the beta model postulated operators that adjust the response strength of each alternative (e.g., $v(x)$) rather than adjusting the choice probabilities. More formally, during a learning experiment on each trial n each alternative $x \in S$ has a response strength $v_n(x)$. If alternative x was selected on trial n then the response strength for trial $n + 1$ is updated. To make a choice on the trial $n + 1$ the individual then uses the choice rule described in Equation 4. Thus, in the beta model, LCA is assumed to hold on any given trial, but between trails the probability changes via adjustments in response strength.

This same basic idea is implemented in recent reinforcement learning algorithms that have become popular in a variety of areas including machine learning, control theory, game theory, and neuroscience. Perhaps the closest relation is with the Q-learning algorithm. According to this algorithm, after an agent selects an alternative x , the agent uses the experienced reward to update an expectation $Q(x)$ for alternative x (Sutton & Barto, 1998). To make a choice between the available alternatives in the set S the algorithm uses the following rule sometimes called softmax

$$P(x) = \frac{exp[Q(x)/\tau]}{\sum_{y \in S} exp[Q(y)/\tau]}. \tag{12}$$

The rule is a particular instance of Equation 4. The parameter τ moderates the degree to which the agent chooses the alternative with the largest expectation. As τ increases, the agent is less likely to choose the option with the largest expectation and instead explore other alternatives. Note unlike the similarity choice model (Equation 8), τ is not specific to the alternative so it characterizes general exploration rather than a bias to a specific alternative. In general, this means the probabilistic choice rule provides a mechanism to tradeoff between selecting an alternative that provides an immediate immediate payoff (exploit) or selecting an alternative that provides information to develop an expectation about future choices (explore). Neuroscientists are beginning to understand the neural basis of these learning mechanisms (Niv, 2009). The softmax expression, in particular, has proved useful in beginning to understand the neural mechanisms underlying the tradeoff between exploration and exploitation (Daw et al., 2006). Finally, these learning mechanisms with the softmax rule have are being integrated into larger cognitive architectures (Fu & Anderson, 2006).

8 Future Directions and Conclusion

LCA challenged conventional theories of choice to move from a deterministic account to a probabilistic one. In so doing, LCA has proved to be a power and useful tool in the modeling toolboxes of social and cognitive scientists for over 50 years. A search among scholarly articles is certain to reveal its vast impact. Yet, there are still areas that are left to be explored with LCA and the accompanying mathematical model of choice. Behaviorally, one area that Luce (1959) discussed and has still remained largely unexplored is in the area of ranking alternatives. Luce (1959) proposed one way to adapt LCA to ranking where the individual essentially makes a series of covert choice first picking say the best alternative from the set, then the second best alternative from remaining items, and so on. This is called forward ranking. Another procedure called backward ranking works the opposite direction picking the worst, next worst and so on. Both procedures seem intuitively plausible; however, forward and backward ranking do not lead to the same solution

(Luce, 1959). For current work in this area see Marley & Louviere (2005) and Lee et al. (2012).

As decision theorists continue to reconcile with the theoretical implications of probabilistic choice behavior there are certainly more and more challenges. One challenge is a psychological one, which is to continue to develop a theory of choice behavior that is grounded in some of the core tenets of LCA (i.e., choice is probabilistic and that there is a relationship in how people choose between contexts), but also reflects other important properties like the context dependence and time course of choice. A second challenge is a neural one, which is to understand the neural circuitry underlying choice behavior (Gold & Shadlen, 2007). A final challenge is a statistical one, which is to develop appropriate statistical methodology to to test probabilistic theories of choice appropriate variability in choice behavior Regenwetter et al. (2011). These are just some of the challenge we face. Regardless, LCA via its simple elegance is certain to continue to be an invaluable tool in helping us understand individual choice behavior.

9 Cross Reference

- behavioral decision research
- bounded rationality
- heuristics for decision and choice
- strategic and decision heuristics
- economic psychology
- paradoxes of choice
- random utility models of choice and response time
- utility and subjective probability
- dynamic decision making
- game theory
- signal detection theory

- stochastic dynamic models
- recognition

References

Adams, E. & Messick, S. (1957). *An axiomatization of Thurstone's successive intervals and paired comparisons scaling models*. Technical report, Stanford University.

Arrow, K. (1951). *Social Choice and Individual Values*. New York: John Wiley and Sons.

Bradley, R. A. & Terry, M. E. (1952). Rank analysis of incomplete block designs. i. the method of paired comparison. *Biometrika*, 39, 324–345.

Clark, F. R. (1957). Constant ratio rule for confusion matrices in speech communication. *Journal of Acoustical Society of America*, 29, 715–720.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095), 876–9.

Debreu, G. (1960). Review of r. d. luce, individual choice behavior: A theoretical analysis. *American Economic Review*, 50, 186–188.

Estes, W. K. (1997). Some reflections on the role of the choice model in theories of categorization, identification, and learning. In A. A. J. Marley (Ed.), *Choice, Decision, and Measurement* (pp. 321–329). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Fu, W. & Anderson, J. A. (2006). From recurrent choice to skill learning: A reinforcement learning model. *Journal of Experimental Psychology: General*, 135, 184–206.

Gold, J. I. & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, 30, 535–574.

Lee, M. D., Steyvers, M., de Young, M., & Miller, B. (2012). Inferring expertise in knowledge and prediction ranking tasks. *Topics in Cognitive Science*, 4, 151–163.

Luce, R. D. (1959). *Individual Choice Behavior*. New York, NY: John Wiley & Sons, Inc.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (pp. 103–189). New York: Wiley.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, 15, 215–233.

Luce, R. D. & Suppes, P. (1965). Preference, utility, and subjective probability. In R. D. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology*, volume 3. New York, NY: Wiley.

Marley, A. A. J. & Louviere, J. J. (2005). Some probabilistic models of best, worst, and best–worst choices. *Journal of Mathematical Psychology*, 49, 464–480.

McFadden, D. (1976). Quantal choice analysis: A survey. *Annals of Economic and Social Measurement*, 5, 363–390.

McFadden, D. (2001). Economic choice. *The Americal Economic Review*, 91(3), 351–378.

McKelvey, R. D. & Palphrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10, 6–38.

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53, 139–154.

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review*, 118(1), 42–56.

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, XLIV, 631–661.

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.

Rumelhart, D. L. & Greeno, J. G. (1971). Similarity between stimuli: An experimental test of the luce and restle choice models. *Journal of Mathematical Psychology*, 8, 370–381.

Shipley, E. F. (1960). A model for detection and recognition with signal uncertainty. *Psychometrika*, 25, 273–289.

Simonson, I. & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, 29(3), 281–295.

Sutton, R. S. & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Tanner, W. P. J. & Swets, J. A. (1954). A decision - making theory of visual detection. *Psychological Review*, 61, 401–409.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273–286.

Train, K. (2003). *Discrete Choice Models with Simulation*. Cambridge: Cambridge University Press.

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

Tversky, A. & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, 101(4), 547–567.

Tversky, A. & Russo, J. E. (1969). Substitutability and similarity in binary choices. *Journal of Mathematical Psychology*, 6(1), 1–12.

Tversky, A. & Simonson, I. (1993). Context-dependent preferences. *Management Science*, 39(10), 1179–1189.

Usher, M. & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, 111(3), 757–69.

Yellott, J. I. (1977). The relationship between luce's choice axiom, thurstone's theory of comparative judgement, and the double exponential distribution. *Journal of Mathematical Psychology*, 15, 109–144.