Process Tracing, Sampling, and Drift Rate Construction

Neil Stewart and Timothy L. Mullett

Warwick Business School, University of Warwick

Author Note

This research was supported by Economic and Social Research Council grants ES/K002201/1, ES/P008976/1, and ES/N018192/1

What can process tracing tell us about the nature of evidence accumulation during choice? In this chapter, we review the idea from a model called decision by sampling (Stewart, Chater, & Brown, 2006; Stewart, Reimers, & Harris, 2015) and show how we have used process tracing data to constrain the development of this model into a process model of choice. In the decision by sampling model decisions are the result of a series of binary, ordinal comparisons between attribute values. That is, people make lots of comparisons between attribute values and use the number of wins each alternative achieves as the criteria for making a decision. Below we explain how people counting wins will behave as if the subjective value for an attribute value is given by its rank position against other attribute values. We discuss how this sort of tally can be particularly robust and can be considered as optimal. We review evidence for the rank hypothesis in judgment and choice, and also evidence from neuroimaging studies including fMRI and single cell recording. We then switch tack and consider the attentional drift diffusion model (Krajbich, Armel, & Rangel, 2010; Krajbich & Rangel, 2011). The attentional drift diffusion model integrates process tracing data from eye movements during choice and attribute values to produce accounts of the choice made, the time taken to make the choice, and the patterns of visual attention seen whilst the choice is made. We conclude with a discussion of how the rank hypothesis and drift diffusion can be combined and developed into a new mathematical model of multiattribute choice, with modeling assumptions constrained by process data.

Decision by Sampling and the Rank Hypothesis in Judgment and Choice

Parducci (1965, 1995) proposed range-frequency theory as an explanation of how people assign a set of categorical judgments to stimuli varying on a single perceptual dimension. For example, people might be assigning a set of category labels “very small”–”very large” to a series of squares varying in their size. Parducci assumes people make these assignments using two principles: the range principle and the frequency principle (see Figure 1). According to the range principle, the stimuli continuum, here square size, is divided into equal width bins. According to the frequency principle, the stimulus continuum is divided up such that each category label is used equally often (e.g., for 3 squares in Figure 1). For example, if people have five categories to use equally often they should put the smallest fifth of squares into the first category, the next smallest fifth into the next category, and so on, no matter what the actual sizes are.

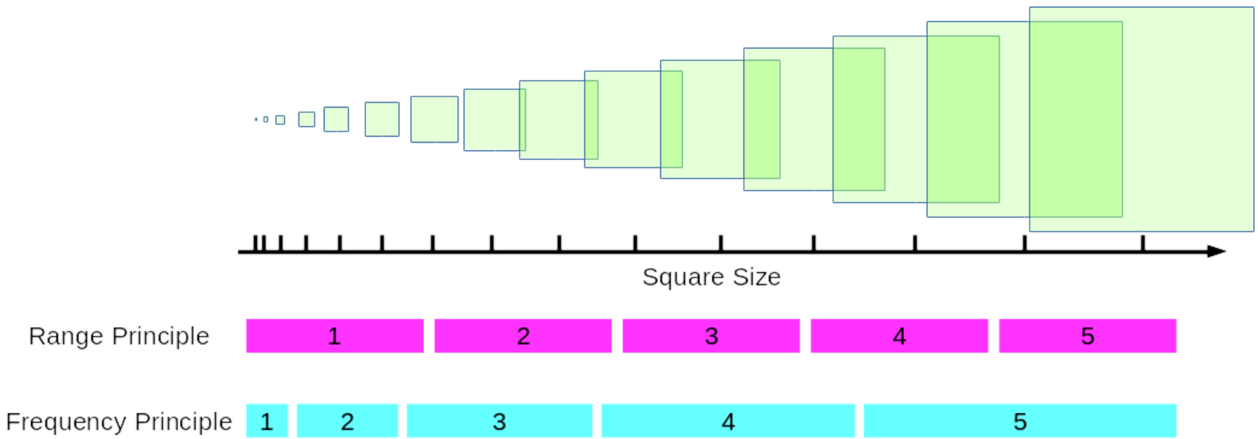


Figure 1. An illustration of how squares are divided into five equal sized bins according to the range principle and five equally used bins according to the frequency principle.

One consequence of the frequency principle is that people will tend to evaluate magnitudes as a function of their ranks. An unpublished experiment from Brown and Saktreger illustrates this well for judgments of the price of bags of sweets varying in weight (though results are similar for perceptual magnitudes, e.g., Parducci & Wedell, 1986). Figure 2 plots the price elicited by an incentive compatible Becker, DeGroot, and Marschak (1964) auction for bags as a function of weight. To facilitate averaging over participants, prices for each participant were divided by the price they judged for the heaviest bag. Brown and Saktreger used two different sets of bags of sweets. In one set the prices of the bags followed a unimodal distribution, with many bags weighing between 100g and 150g (see the rug plot at the top of Figure 2). In the other set the bags followed a bimodal distribution, with many bags less than 100g and may bags more than 150g (see the rug at the bottom of Figure 2). Brown and Saktreger were careful to hold the range of weights fixed across the sets, and to include common weights of 100g, 125g, and 150g in both sets. Prices are very different for the 100g and 150g bags across the sets as this is where their respective rank positions within the sets differ most. The dashed lines are the predictions of the frequency principle, and match the qualitative pattern in the data well. Both predictions and observed judgments increase most quickly where the distribution is most dense, as this is where rank position changes most quickly.

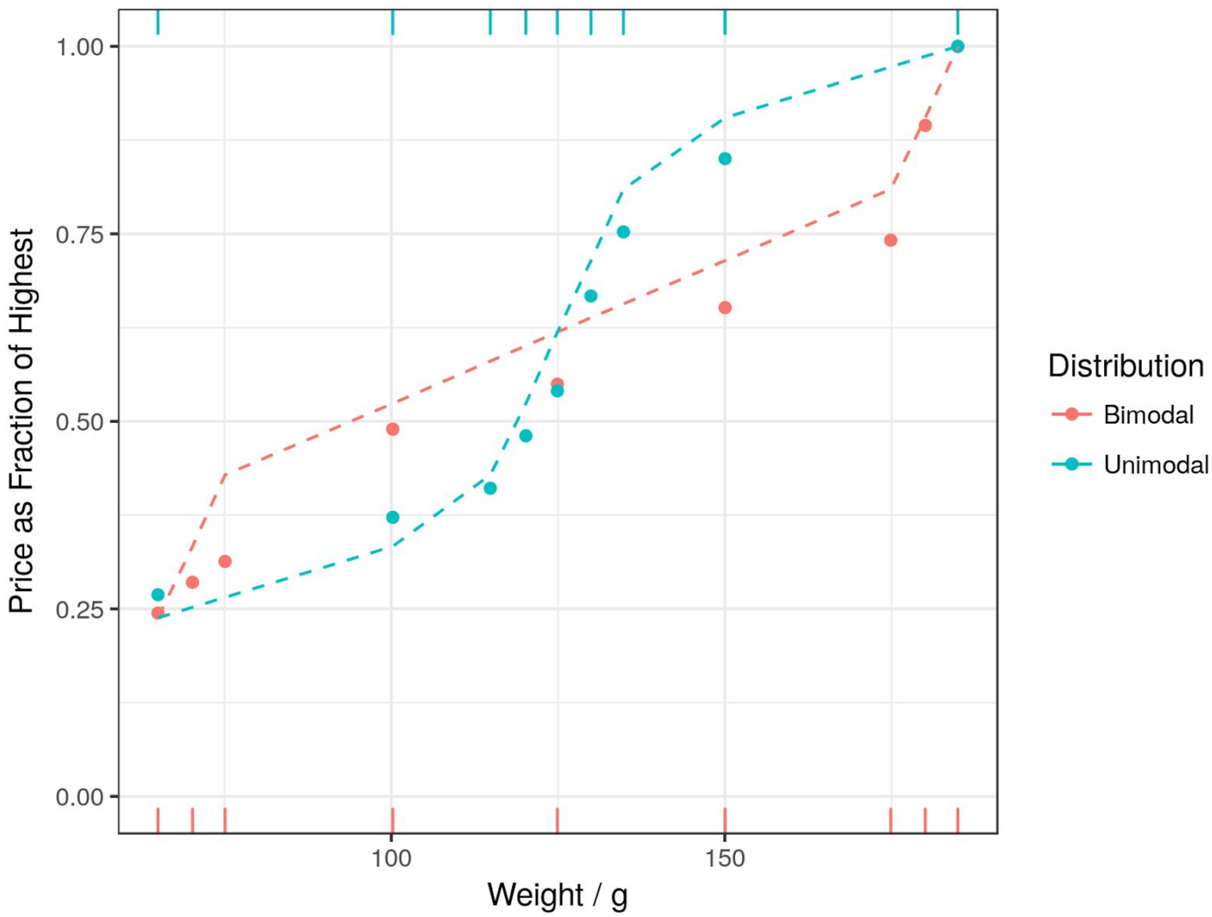


Figure 2. Judged price for a bag of sweets as a function of the bag's weight for two different sets of bags. Dashed lines are the predictions of the frequency principle.

Drawing upon Parducci's frequency principle, the original exposition of decision by sampling model (Stewart et al., 2006) offers an account of the origin of the subjective values for economic quantities like money, risk, and time. Consider a person forming a subjective valuation for £100. The idea is that they compare the £100 against a series of sums of money sampled from their memories—perhaps other values presented in the choice context, or values that come to mind because they have been recently experienced or are evoked by the context. Suppose, in this example, people have in mind the sums £10, £17, £30, and £240. £100 compares favorably with £10, £17, and £30, and unfavorably with £240. This three of the four possible comparisons with the sample in memory would lead to a win for £100. The probability that each comparison leads to a win for £100 is 3/4, which means people would behave as if, in this context, £100 has a subjective value of 3/4.

Chapter [Wulff chapter] provides a discussion on process of external information search in decision making, which is one origin for the samples people might have in mind. Though information search is clearly critical, Stewart et al. (2006) sidestepped detailed assumptions about information search by assuming that the distributions people have in mind reflect the distributions of attribute values in the environment (cf Anderson & Schooler, 1991). Stewart et al. used credits and debits into bank accounts as a proxy for the shapes of the distributions of gains and losses that people experience in the environment. Figure 3 shows the frequencies of different credits (left) and debits (middle) in a large sample of bank accounts. Both distributions have the property that there are more small transactions than large ones, and the (very approximate) straight line in log-log space indicates the distributions follow something resembling a power law. Figure 3 also shows the resulting decision by sampling value function (right). The top right quadrant plots the relative rank of a given gain as a function of the size of the gain. The function is steep just above zero, because that is where the distribution of credits is most dense—and thus where a given absolute increase in the size of a gain has the largest effect on rank position. But nearer £1,000 the function is flat, because that is where the distribution of credits is less dense—and thus where a given absolute increment in the size of a gain has only a small effect on rank position. The bottom left quadrant plots the value function for losses. The function is plotted upside down because large losses are bad (whereas large gains are good). Like the function for gains, the function is steepest near 0 where the debits are most dense and flatter for higher magnitude losses where the debits are much less dense. The asymmetry in the function, with a steeper initial function for losses than gains results from the asymmetry in the distribution of gains and losses in the credits and debits, where there are more small debits than small credits. This decision by sampling value function bears a striking resemblance to the value function from prospect theory (Kahneman & Tversky, 1979). In prospect theory the shape of the value function is motivated by risky choice phenomena, including risk aversion for gambles involving gains and risk seeking for gambles involving losses. In decision by sampling, the shape emerges from the environmental distribution of gains and losses combined with the cognitive process of counting favorable binary comparisons. In other words, decision by sampling offers an explanation, in terms of the distribution of values in the environment, as to why people make the decisions they do, whereas prospect theory just describes the decisions.

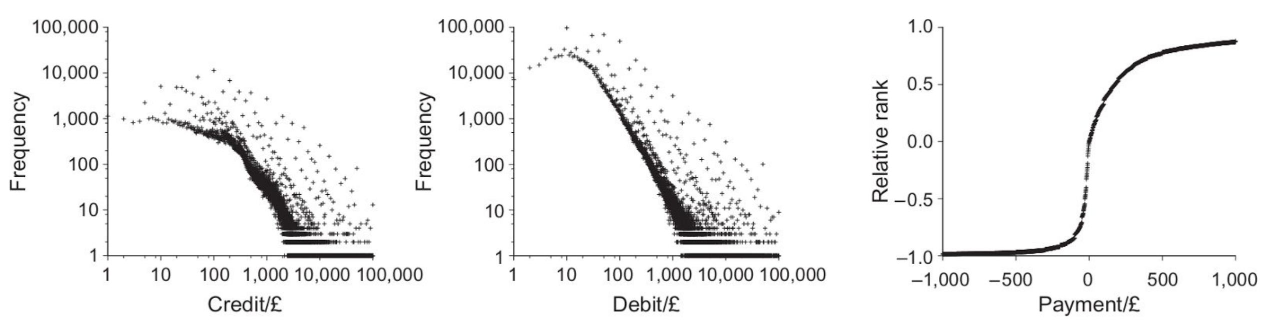


Figure 3. The frequencies of credits and debits in bank accounts (left and middle) and the value function that would result from sampling from these credits and debits (right) (after Stewart, Chater, & Brown, 2006, Figures 1 and 2, based on a sample of 1,800,000 transactions).

Decision by sampling is able to offer an account of the risky choice phenomena that motivated prospect theory because it predicts that people will behave as if they have a prospect theory value function in worlds with positively skewed distributions of gains and losses. Stewart and Simpson (2008) and Stewart (2009) developed the model from judgments into choice. They proposed an explicit mechanism where by wins for each alternative are accumulated until a threshold is reached, with the probability of a win for each attribute of each alternative defined just as above. The model is able, for example, to explain the patterns in risky choice seen in Kahneman and Tversky (1979) that were used to motivate prospect theory, like, for example, the Allais paradox. In the Allais paradox people prefer a sure $3,000 over a 75% chance of $4,000 otherwise nothing, but prefer a 20% chance of $4,000 over a 25% chance of $3,000. This is a violation of expected utility because people reverse their preferences for the $3,000 option over the $4,000 option across the pair of choices even though the first choice is related to the second choice simply by dividing the probabilities by 4. In decision by sampling, the reversal occurs because the difference in the relative ranks of 75% and 100% is much larger than the difference in the relative ranks of 20% and 25% within the sample of probabilities people have in mind (Stewart & Simpson, 2008).

Ungemach, Stewart, and Chater (2011) conducted a test of this model. According to decision by sampling, changes to the amount of money available in people's memory should change their subjective valuations of those amounts. To test this, Ungemach et al. used visits to the campus supermarket as a natural priming experiment. Upon leaving the store, people were asked to give up their receipt in return for a choice between two lotteries. One lottery offering an unlikely prize of £1.50 and the other offered a more likely prize of £0.50. Thus participants faced a choice between a higher risk but higher reward lottery and a lower risk but lower reward lottery. Ungemach et al. found that the fraction of item prices on the supermarket receipt between £0.50 and £1.50 predicted the lottery choice. Assuming that the prices of the supermarket receipts have some correspondence to the prices in memory, decision by sampling makes a clear prediction. When most of the items were between £0.50 and £1.50, £0.50 is one of the lowest ranking amounts in memory and £1.50 is one of the highest ranking amounts in memory. But when most of the items were lower than £0.50 or higher than £1.50, £0.50 and £1.50 will get very similar ranks. The proportion of prices on the bill between £0.50 and £1.50 did indeed predict preference for the high risk £1.50 lottery, while the total spend, and other economic variables, did not (see Matthews, 2012, for a failure to replicate in the intertemporal domain and Canic, 2016, for a meta analysis of studies with the same logic in choices between simple lotteries).

The idea in decision by sampling, that values are judged against a small sample of comparison items, is related to ideas in norm theory (Kahneman & Miller, 1986) where the typicality or normality of a hypothesis is judged against a sample of counterfactual hypotheses evoked by the context, and support theory (Tversky & Koehler, 1994) where the likelihood of a hypothesis is based upon comparisons to a sample of alternative hypotheses. It is also related the notion of the construction of preference (Bettman, Luce, & Payne, 1998; Payne, Bettman, & Johnson, 1992; Slovic, 1995), because the subjective value of attribute values is derived on the fly during the series of binary ordinal comparisons, rather than being static. Finally, the counting of favorable binary ordinal comparisons is tallying (Dawes, 1979), which is quite robust because less information needs to be estimated about how to combine information across dimensions (signs of coefficients are sufficient, rather than magnitudes). This is because tallying is formally equivalent to a ridge regression with a strong penalty term which is formally equivalent to a Bayesian prior upon that regression weights are near zero (Parpart, Jones, & Love, 2017), a prior that happens to coincide with a truth about the environment.

Diffusion Models

One weakness of decision by sampling as a process model is that it makes only coarse predictions for reaction time. In cases where there are relatively few binary comparisons to be made, decision by sampling also predicts that the current estimate of value may undergo large changes during deliberation. This is because when there are few possible comparisons, any single one will be a large proportion of the underlying value information. This suggests that any process measure of the current value estimate should also show this coarse step change in the representation of value. This is important because neuroscience methods have provided insights into the moment to moment representation of value at a neural level, and how this information is combined to reach a decision. The findings suggest that rather than exhibiting dramatic changes during deliberation, the neural representation of value is noisy moment-to-moment, but on average it is relatively stable (Thorpe, Rolls & Maddison, 1983; Tremblay and Schultz, 1999).

Some alternative models do much better at predicting both the neuroimaging patterns and reaction times. In particular, those predicated upon evidence accumulation, such as drift diffusion. This family of models assumes that because the neural representation of value is so noisy moment-to-moment, the decision maker accumulates evidence. This is akin to drawing repeated samples from this noisy neural representation of value so that the average can be estimated. Within the drift diffusion framework this is thought of within the structure of a single number: the accumulator. The value of option A has a positive effect on this number, and the value of option B has a negative effect on this number. Multiple samples are taken from the noisy value representations meaning that, on average, if option A has a higher value, the accumulator will become more positive over time, and if option B is more valuable, the accumulator will become more negative over time. This is referred to as drift, because over time the accumulator number will drift higher or drift lower. This will continue until the accumulator value becomes so positive or so negative that it hits a positive or negative decision threshold and thus the relevant option is selected. Because this process is assumed to be noisy, these models do well at predicting reliable patterns in choices and response times, such as easier choices being quicker, and that response times will exhibit a positive skew.

There are several successful models that make use of the drift diffusion mechanism (Busemeyer & Townsend, 1993; Roe, Busemeyer & Townsend, 2001; Ratcliff, 1978; Ratcliff & Rouder, 1998). In addition to having high accuracy when predicting choice error rates and reaction times, they have a degree of neural plausibility that few other approaches can match (Gold & Shadlen, 2007); a noisy accumulation system can be efficiently performed by neuronal firing patterns, and structures necessary for the implementation of choice have been identified in the brain. They can also predict the relative stability of average neural activity in reward regions by assuming this activity is a temporally smoothed representation of drift rate, and some degree of context dependency by assuming interactions and inhibitory links between attributes and options (Busemeyer, Jessup, Johnson, & Townsend, 2006). More recent developments of this class of models also incorporate findings from eye tracking and attention research (Krajbich et al., 2010). By assuming that evidence accumulation is biased to accumulate evidence in favour of the option currently being looked at, the model is able to fit numerous aspects of visual attention during deliberation, as well as improving predictive accuracy in other measures by incorporating attention as a predictor.

Despite their success, there are some fundamental questions that evidence accumulation models struggle to answer. One of the most fundamental of these is where does the initial representation of value come from? These accumulation models do very well at describing how this noisy signal is accumulated and turned into decisions and reaction times, but are silent on how the properties of the stimuli are turned into this noisy signal. The most common assumption is that the drift rate is simply a linear transformation of the objective stimulus value, sometimes with additional tweaks and biases that allows the model to capture non-linear patterns of comparison or biases to give some attributes more weight than others. However, even the more complex assumptions still fail to capture many contextual phenomena that are quite naturally explained by decision by sampling.

A question therefore becomes, could a sampling and comparison process play a role in the construction of these noisy representations of value? One way to answer this is to turn to evidence from neuroscience. These techniques allows for a direct measurement of activity in regions known to represent option values, and the drift rate component of evidence accumulation models (for a review see Konovalov & Krajbich, 2016). One of the key areas for value representation is a frontal region called the vmPFC (ventro-medial Pre-Frontal Cortex; Plassmann, O'Doherty, & Rangel, 2007; Clithero & Rangel, 2014). Results from fMRI experiments show that there is greater activity here when an individual is considering a more valuable item. This activity is also relatively stable on average, rather than representing an increasing activity pattern as more evidence is accumulated, so we can be relatively confident that it is representing the value signal/drift rate (for discussion of which candidate regions may be encoding the decision or evidence accumulation see Mazurek et al., 2003; and Katz et al., 2016).

By looking at activity in the vmPFC we can therefore test different assumptions of how this value signal is constructed. A simple way to do this is with money, as different amounts of money can be easily compared on the same scale. However, the numerical properties of money are difficult to represent neurally: financial reward can increase infinitely, but neural activity (neuron firing rates) has a finite range. This suggests that in a given task, there must be some mechanism that can rescale to the range of values on offer. The most common assumption is that neural firing linearly rescales based upon the range of values in the environment (Rangel & Clithero, 2012, and cf. Parducci's range principle) or upon the mean of the values in the environment (Knutson & Peterson, 2005; and cf. Helson's, 1964, adaptation level principle). However, this mechanism does not answer how value is perceived and constructed in the first place, before undergoing this transformation. In fact it assumes even more representation of absolute values - for the range or mean of values - which simply moves the problem of magnitude representation to a different part of the process. Furthermore, no neural regions have been identified that represent range or averages.

If drift rates are constructed by sampling, many of these issues are sidestepped. If decisions, and thus drift rates, are based upon some mechanism of sampling and comparison then such a mechanism would inherently normalise the encoding of value. This is because, it is likely that the alternatives sampled will be relevant to the current choice and therefore on a similar scale to the currently attended item; though even if they are very different magnitudes, an ordinal comparison is still simple. Value can then be encoded entirely in terms of favorable/unfavorable comparisons. This is easy to test because it makes unique predictions that cannot be explained by other rescaling hypotheses. Specifically, the decision by sampling framework of ordinal comparisons predicts that the neural activity representing value will be based upon rank. This leads to unique predictions when the distribution of values in the environment is non-linear or unevenly distributed in some way. When Mullett and Tunney (2013) presented values to subjects that were drawn from a very non-linear environment, it was found that activity in dopaminergic regions of the ventral striatum, and the vmPFC represented the reward's rank position within the distribution of values. The effect was sufficiently strong that the difference in fMRI neural signal (BOLD) between 10p and 30p stimuli was the same as the difference between £5 and £10 despite the difference in values being more than 16 times larger in the latter case.

A sampling based mechanism would also explain the robust within choice context effects showing that the neural activity in the vmPFC is encoded based on whether the current option compares relatively favorably or unfavorably with the alternative outcomes on a given trial. For example, when a monkey is told they will be awarded a piece of apple, neural firing rates are high if the other possible outcome was a piece of (less preferred) cereal, but low if the other possible outcome was a piece of (more preferred) raisin (Tremblay and Schultz, 1999). A similar effect is found in humans when receiving monetary rewards: neural firing for a win of 50p is high when the alternative possibility was 10p, and low when the alternative possibility was £1 (Elliott et al., 2008). This is also true when individuals switch their attention during deliberation, with neural firing being higher whilst they are looking at a higher value reward (Lim, O'Doherty, & Rangel, 2011; McGinty, Rangel & Newsome, 2016).

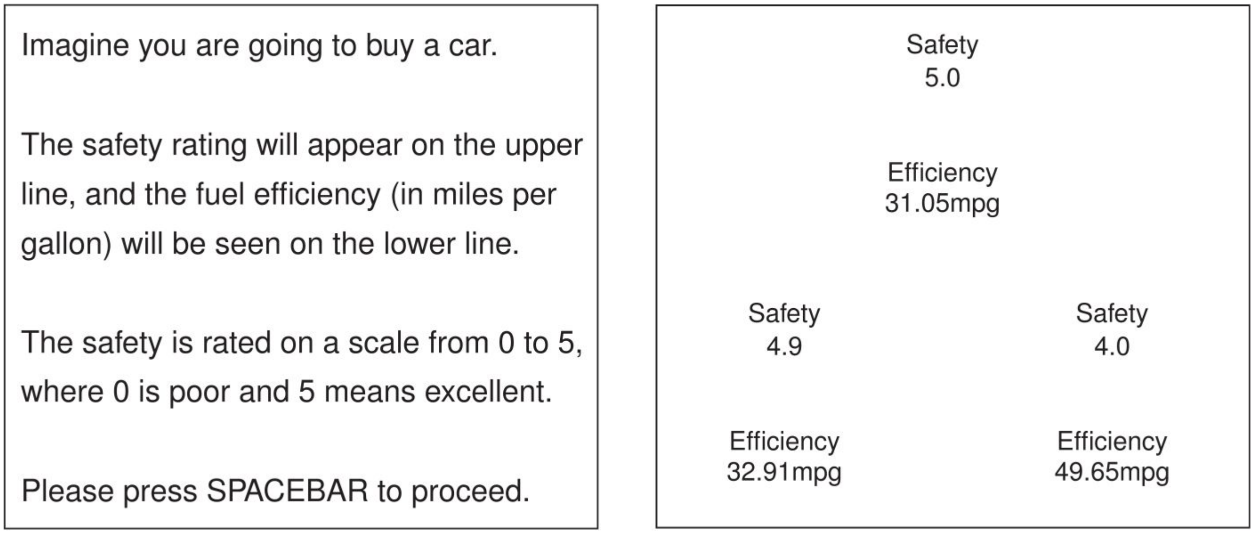
To this point, we have reviewed the decision by sampling model and some of the context effects in studies of economic judgment and choice that supports the assumption that judgment and choice are sensitive to rank effects. We have also reviewed evidence from neuroeconomics suggesting that evidence is accumulated over time until a choice is made, a process well accounted for by drift diffusion models. These models, however, do not offer a natural account of where drift rates come from and how they might vary across different contexts. Below we review how process tracing evidence has been used to constrain the integration of the decision by sampling model and the drift diffusion / evidence accumulation frameworks.

Using Process Tracing to Constrain Assumptions: Multialternative Decision by Sampling

Below we review key evidence from process-tracing studies which we have used to constrain ideas about how evidence is aggregated over time in the accumulator framework. Both sources of evidence involve eye movements (see Chapters 3–6): The first explores the link between eye movements between attribute values in multialternative choice and the choice ultimately made. The second concerns changes in the pattern of eye movements in the run up to a choice.

Alternatives are compared in pairwise comparisons on single dimensions

Noguchi and Stewart (2014) explored the link between eye movements and choice, testing in particular the idea that the accumulation of favorable binary, ordinal comparisons between pairs of attribute values might be driving choice. Participants made a series of choices, each between three alternatives on two dimensions. For example, one choice was between three cars, each with different safety ratings and fuel efficiency. Figure 4 shows two screenshots from the experiment for the car choice. Each choice involved a different consumer-good cover story.



(a) Choice Description (b) Choice Set

Figure 4. A screenshot from Noguchi and Stewart (2014, their Figure 3), depicting a choice between three cars.

The attribute values were chosen to create attraction, similarity, and compromise questions. The attraction, similarity, and compromise effects are important in multialternative choice because together they set a significant benchmark for modelling (see Roe et al., 2001). Figure 5 illustrates the choice sets used to demonstrate the attraction, similarity, and compromise effects. The alternatives vary on two attribute dimensions *x* and *y*, with more of each dimension being better. Each set starts with a pair of core alternatives, A and B, which are, roughly, equally likely to be chosen. Adding either D, S, or C to the choice set, leads to a favoring of alternative A over alternative B. In particular, D is asymmetrically dominated by A, adding D to the choice set increases the choice share for A over B due to the attraction effect. Adding alternative S, which is similar to B, also leads to an increase preference for A over B—the similarity effect. Adding a competitive but distant option C makes A a compromise between C and B and increases the relative preference for A over B. The attraction, similarity, and compromise effects represent a violation of regularity (the property that adding an alternative cannot increase the choice share for any existing alternative) and independence from irrelevant alternatives (the property that adding an alternative cannot reverse the ordering of choice shares for existing options). Both of these principles are properties of any model in which the value of an alternative remains stable across different choices and is invariant to the other alternatives in the choice set (see Rieskamp, Busemeyer, & Mellers, 2006, for a review). As such, the existence of the attraction, similarity, and compromise effects points towards a construction of preference, where people behave as if the values of the different alternatives are constructed during the choice process in a context-dependant way.

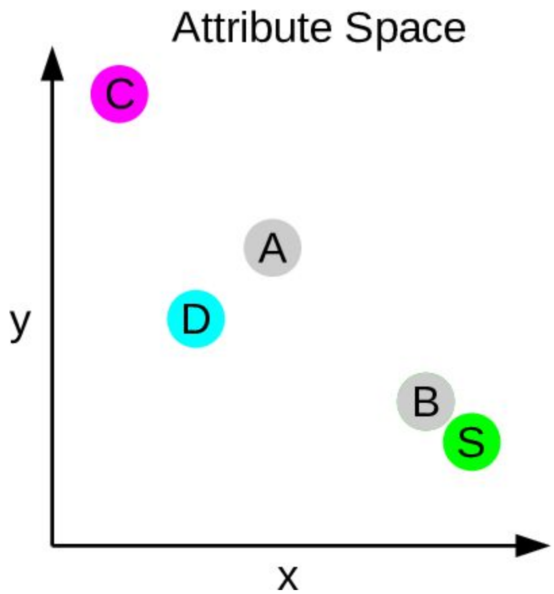


Figure 5. The attraction, similarity, and compromise questions sets. A and B are common alternatives. The addition of an asymmetrically dominated alternative D creates an attraction choice set favoring A, the addition of a similar alternative S creates a similarity choice set favoring A, and the addition of a distant alternative C creates a compromise choice set favoring A.

The models of these effects are all quite similar, in that they all embody the idea that microsamples of evidence are accumulated over time until a decision is reached (e.g., Bhatia, 2013; Roe et al. 2001; Trueblood, Brown, & Heathcote, 2014; Usher & McClelland, 2004; Wollschläger & Diederich, 2012). Where they differ is in how they define the evidence that is accumulated—and it is this question that Noguchi and Stewart were attempting to answer using process tracing data. Noguchi and Stewart recorded participants eye movements while they made each choice. They considered transitions between attribute values—defined as moving gaze from one attribute to another. Each transition can be defined as favoring one alternative. For example, the transition between Alternatives A and B on Dimension X favors Alternative B (because Alternative B is better on Dimension X) and the transition between Alternatives A and B on Dimension Y favors Alternative A (because Alternative A is better on Dimension Y). For each choice, Noguchi and Stewart counted the number of transitions favoring each alternative. Figure 6 shows how the number of transitions favoring an option is higher on trials where that option was indeed chosen. That is, the trials where Alternative A is chosen are those where transitions that favor Alternative A happen to have been made more often and transitions favoring the other options were made less often. For example, for the attraction choices in Figure 6A, prior to choosing A (○ symbols) the transitions which favor A are more frequent and the transitions favoring B and D are less frequent. Prior to choosing B (■ symbols) the pattern reverses, so that transitions favoring A are less frequent and transitions favoring B are more frequent. And thus these fluctuations over trials in the pattern of eye movements are associated with changes in the alternative finally chosen.

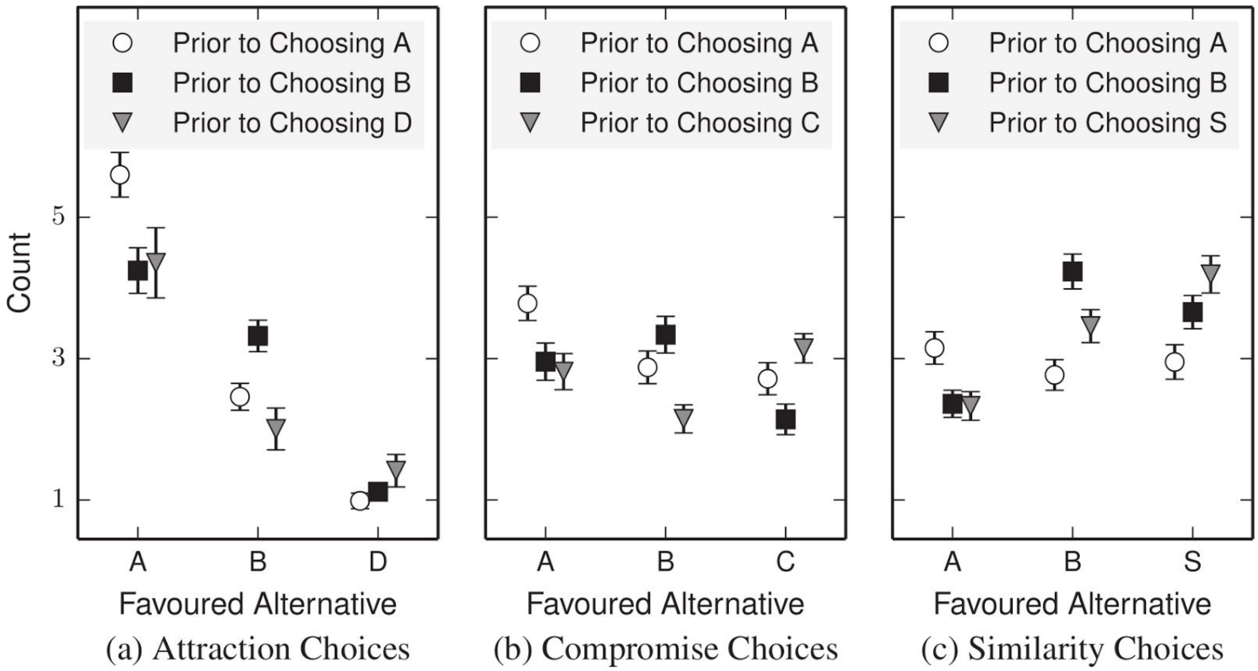


Figure 6. The number of transitions made during a choice favoring each alternative, separately for each choice outcome (from Noguchi & Stewart, 2014, Figure 11).

Noguchi and Stewart (2014) used these results to discriminate between different types of evidence accumulation. The finding above is most consistent with *attribute-and-alternative-wise* models in which one pair of alternatives are compared on a single dimension at each point in the accumulation process. *Alternative-wise* models assume that all of the attributes of an alternative are integrated before comparisons. *Attribute-wise* models assume that one dimension is attended and all of the alternatives (not just the two in the attended pair) are simultaneously evaluated on that dimension. Neither attribute-wise models nor alternative wise models can predict the interaction between the pair of alternatives and the dimension attended. So, because the process data implicate an attribute-and-alternative-wise model, multialternative decision by sampling was implemented as an attribute-and-alternative-wise model. The process data constrained the modeling assumption.

The gaze bias effect implicates a relative stopping rule

One of the most robust findings in visual attention during choice is the late onset bias (also known as the gaze cascade). This describes the phenomenon that in the final moments before a choice is made, subjects become increasingly likely to attend the option they subsequently choose (Figure 7). In a series of model simulations Mullett and Stewart (2016) show that this pattern can be captured by an evidence accumulation model with a simple attentional bias assumption that is present in a number of existing models: evidence is accumulated more rapidly for the option that is currently attended (e.g., Krajbich et al., 2010). However, these models can only capture the effect if they also assume a relative decision threshold, that is, a decision is made once one of the options has accumulated a total amount of evidence that is X *more than* the evidence for the other alternative. This is in contrast to models which assume an absolute threshold, where a decision is made once any one option has accumulated an amount of evidence that surpasses a pre-defined amount of X, regardless of whether the amount accumulated for the alternative is very similar, or much lower. The underlying conceptual cause is that in a relative threshold model, whenever evidence is accumulated for one of the options, it changes the relative amount of evidence in that options favor, i.e. closer towards that threshold. However, the relative nature means that it also moves further away from the decision threshold for the competitor. Because of this back and forth nature of a relative rule, it is necessary for there to be a run of evidence accumulated in favor of an item. Without this run, or series of samples in favor of the option, the relative evidence would not stray over the threshold before the bias changed to favor the alternative. This need for a run of samples prior to crossing a threshold means that the bias in sampling must onset prior to choice. Furthermore, the bias will appear to develop smoothly when averaged across choices because the random noise in evidence accumulation means that sometimes the accumulator will already be close to a threshold and only require a relatively short run, whereas other times it will be further away and require a longer run (which is less likely to occur due to the variable distribution of attention, and thus “run length”).

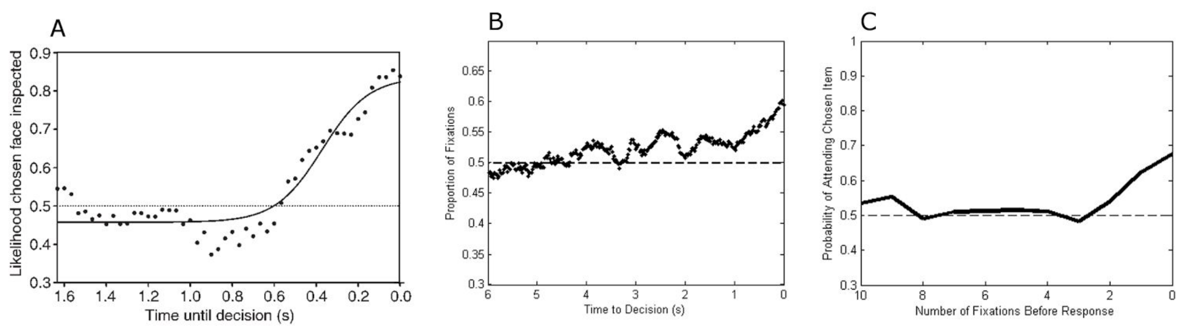


Figure 7. The late onset bias as measured in: (A) a face preference task (Shimojo, Simion, & Shimojo, 2003); (B) a multi-attribute choice task (Mullett & Tunney, 2015); (C) risky gamble choices (Stewart, Hermans, & Matthews, 2015). Adapted from Mullett and Stewart (2016).

This finding demonstrates the importance of process data in constraining models. In any model where there is an increasing bias towards accumulating evidence or sampling in favor of the attended item, the decision rule must be relative. Only a relative model involves a run of evidence for one alternative in the run up to a decision. In fact, the late onset bias cannot be captured in an absolute model even when assuming a feedback loop between evidence accumulated and attention. For this reason, the decision rule in multialternative decision by sampling is a relative decision rule. Again, the process data constrained the modeling assumption.

Multialternative Decision by Sampling

The process tracing evidence from Noguchi and Stewart (2014) suggest that alternatives are compared in pairs on single dimensions, at least in attraction, similarity, and compromise-like multialternative choice sets. The Mullett and Stewart (2016) simulations implicate a relative stopping rule. Noguchi and Stewart (2017) have used these two pieces of process tracing evidence to constrain a new accumulator model called multialternative decision by sampling. The models offers a quantitative account of the attraction, similarity, and compromise effects equal to that of competitor models like decision field theory (Roe et al. 2001) and multialternative linear ballistic accumulators (Trueblood et al, 2014), and, at the same time, provides the most broad account of the range of qualitative phenomena in the multialternative choice based on a review of the literature, including the location of decoys, time pressure effects, alignability effects, attribute range and spacing effects, background contrast effects, less is more effects, perceptual focus effects and phantom decoy effects (Noguchi & Stewart, 2017). We think that constraining the modeling assumptions with process data has contributed to the success of the model in explaining the wide range of multialternative choice phenomena.

References

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2,* 396–408. doi:10.1111/j.1467-9280.1991.tb00174.x

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science, 9,* 226–232.

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research, 25,* 187–217.

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review, 120,* 522–543. doi:10.1037/a0032457

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100,* 432–459. doi:10.1037/0033-295X.100.3.432

Busemeyer, J. R., Jessup, R. K., Johnson, J. G., & Townsend, J. T. (2006). Building bridges between neural models and complex decision making behaviour. *Neural Networks, 19,* 1047–1058.

Canic, E. (2016). *Value is context dependent: On comparison processes and rank order in choice and judgment* (Unpublished doctoral dissertation). University of Warwick, Coventry, England.

Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *Journal of Neuroscience, 29,* 12315–12320.

Clithero, J. A., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience, 9,* 1289–1302.

Dawes, R. M. (1979). The robust beauty of linear models of decision making. *American Psychologist, 34,* 571–582.

Elliott, R., Agnew, Z., & Deakin, J. F. W. (2008). Medial orbitofrontal cortex codes relative rather than absolute value of financial rewards in humans. *European Journal of Neuroscience, 27,* 2213–2218.

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

Hare, T. A., Camerer, C. F., Knoepfle, D. T., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal of Neuroscience, 30,* 583–590.

Helson, H. (1964). *Adaptation-level theory.* New York: Harper & Row.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153. doi:10.1037/0033-295X.93.2.136

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291. doi:10.2307/1914185

Katz, L. N., Yates, J. L., Pillow, J. W., & Huk, A. C. (2016). Dissociated functional significance of decision-related activity in the primate dorsal stream. *Nature, 535,* 285–8.

Knutson, B., & Peterson, R. (2005). Neurally reconstructing expected utility. *Games and Economic Behavior, 52,* 305–315.

Konovalov, A., & Krajbich, I. (2016). Over a decade of neuroeconomics: What have we learned? *Organizational Research Methods.*

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience, 13,* 1292–1298. doi:10.1038/nn.2635

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 13852–13857. doi:10.1073/pnas.1101328108

Kvam, P. D., Pleskac, T. J., Yu, S. L., & Busemeyer, J. R. (2015). Interference effects of choice on confidence: Quantum characteristics of evidence accumulation. *Proceedings of the National Academy of Sciences of the United States of America, 112,* 10645–10650. doi:10.1073/pnas.1500688112

Matthews, W. J. (2012). How much do incidental values affect the judgment of time? *Psychological Science, 23,* 1432–1434. doi:10.1177/0956797612441609

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral Cortex, 13,* 1257–1269.

McGinty, V. B., Rangel, A., & Newsome, W. T. (2016). Orbitofrontal cortex value signals depend on fixation location during free viewing. *Neuron, 90,* 1299–1311.

Mullett, T. L., & Stewart, N. (2016). Implications of visual attention phenomena for models of preferential choice. *Decision, 3,* 231–253. doi:10.1037/dec0000049

Mullett, T. L., & Tunney, R. J. (2013). Value representations by rank order in a distributed network of varying context dependency. *Brain and Cognition, 82,* 76–83. doi:10.1016/j.bandc.2013.02.010

Mullett, T. L., & Tunney, R. J. (2015). Attention in multi-attribute choice. *Working paper*

Noguchi, T., & Stewart, N. (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition, 132,* 44–56. doi:10.1016/j.cognition.2014.03.006

Noguchi, T., & Stewart, N. (2017). *Multialternative decision by sampling.* Manuscript submitted for publication.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review, 72,* 407–418.

Parducci, A. (1995). *Happiness, pleasure and judgment: The contextual theory and its applications.* Mahwah, NJ: Erlbaum.

Parducci, A., & Wedell, D. H. (1986). The category effect with rating-scales: Number of categories, number of stimuli, and method of presentation. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 496-516.

Parpart, P., Jones, M., & Love, B. C. (2017). *Heuristics as Bayesian inference under extreme priors.* Manuscript submitted for publication.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology, 43,* 87–131.

Rangel, A., & Clithero, J. A. (2012). Value normalization in decision making: theory and evidence. *Current Opinion in Neurobiology, 22,* 970–81.

Ratcliff, R. (1978). Theory of memory retrieval. *Psychological Review, 85,* 59–108. doi:10.1037/0033-295X.85.2.59

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9,* 347–356. doi:10.1111/1467-9280.00067

Rieskamp, J., Busemeyer, J. R., & Mellers, B. (2006). Extending the bounds of rationality: A review of research on preferential choice. *Journal of Economic Literature, 44,* 631–661. doi:10.1257/jel.44.3.631

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review, 108,* 370–392. doi:10.1037/0033-295X.108.2.370

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience, 6,* 1317–1322.

Slovic, P. (1995). The construction of preference. *American Psychologist, 50,* 364–371.

Stewart, N. (2009). Decision by sampling: The role of the decision environment in risky choice. *Quarterly Journal of Experimental Psychology, 62,* 1041–1062. doi:10.1080/17470210902747112

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology, 53,* 1–26. doi:10.1016/j.cogpsych.2005.10.003

Stewart, N., Hermens, F., & Matthews, W. J. (2016). Eye movements in risky choice. *Journal of Behavioral Decision Making, 29,* 116–136.

Stewart, N., Reimers, S., & Harris, A. J. L. (2015). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. *Management Science, 61,* 687–705. doi:10.1287/mnsc.2013.1853

Stewart, N., & Simpson, K. (2008). A decision-by-sampling account of decision under risk. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 261–276). Oxford, England: Oxford University Press.

Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: From independence to competition. *Psychological Review, 120,* 1–38. doi:10.1037/a0030776

Thorpe, S. J., Rolls, E. T., & Maddison, S. (1983). The orbitofrontal cortex: neuronal activity in the behaving monkey. *Experimental Brain Research, 49,* 93–115.

Tremblay, L., & Schultz, W. (1999, April 22). Relative reward preference in primate orbitofrontal cortex. *Nature, 398,* 704–708. doi:10.1038/19525

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. Psychological Review, 121, 179–205. doi:10.1037/a0036137

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

Ungemach, C., Stewart, N., & Reimers, S. (2011). How incidental values from our environment affect decisions about money, risk, and delay. *Psychological Science, 22,* 253–260. doi:10.1177/0956797610396225

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multi-alternative choice. *Psychological Review, 111,* 757–769. doi:10.1037/0033-295X.111.3.757

Wollschläger, L. M., & Diederich, A. (2012). The 2N-ary choice tree model for N-alternative preferential choice. *Frontiers In Psychology, 3,* 1–11. doi:10.3389/fpsyg.2012.00189