**Neurocomputational Modelling of Altruistic Behavior in Contexts of Loss, Familiarity, and Uncertainty Using Drift Diffusion Models**

Tomas D. Winegar and Parita M. Shah

Department of Neuroscience, Harvard University

NEURO 1401: Computational Cognitive Neuroscience

Dr. Samuel J. Gershman

May 5, 2022

**Abstract**

We aimed to expand the neurocomputational model developed by Hutcherson et al (2015) to accurately model human altruistic behavior in contexts of loss, familiarity, and uncertainty. We made two distinct modifications to random decision variable computation: a social distance factor > 1, and a parameter representing probability of an outcome happening. We also applied the model to a dataset containing values framed as losses. We computed generosity rates and response times and our model provided data which was supported by behavioral findings for each of our extensions. In the loss domain we found that response times and acceptance rates for altruistic decisions were lower. In the context of familiarity, we found that decreased social distance lead to more altruistic choices in the gain domain and less risk taking in the uncertainty domain. For uncertainty in altruistic decisions, we found that response times and acceptance rates were lower than for certain decisions. Therefore, our model highlights how computations of choice may explain the altruistic behavior patterns shown by humans in contexts of loss, familiarity, and uncertainty.

**Introduction**

Many species have the instinctive need to help other individuals. Monkeys will help let another monkey out of a cage in order to share food, rats will help other rats who look like they are drowning, and pigeons will protect a flock eating from predator hawks (de Waal et al., 2008; Roeder, 2013; Sato et al., 2015). Humans also show this type of behavior on varying levels from helping beggars on the street to heroic acts such as saving children who have fallen into a rivers (Sanders & Tamma, 2015; Stromberg, 2014). In all of these cases, on the surface level there is little to no benefit to the individuals helping the stranger, and in many cases, helping directly leads to a disadvantage when compared to not helping, either by losing access to some resources or putting oneself directly in danger. The behavior described here is known as altruistic behavior and is defined as “behavior that benefits another individual at a cost to oneself” by the American Psychological Association (2015).

Questions about altruistic behavior that interest researchers are how, why and through what neurological processes altruism has evolved in different species. In a broad sense, the existence of altruism in many animal species does not seem to make sense in terms of natural selection. If anything, hypotheses would predict that increased survival, and therefore selection, would occur in populations who exhibited opposite, selfish behavior.

**Evolution of Altruistic Behavior in Animals**

Researchers have come up with three main theories as to how altruistic behavior can be explained from an evolutionary perspective: direct reciprocity, kin selection and inclusive fitness (Fehr & Fischbacher, 2003; Kay et al., 2020; Wang & Lu, 2018). The theory of direct reciprocity suggests that altruistic behavior evolved due to the fact that when an individual helps others that are in need, they are more likely to receive the same help back when they, themselves, are in need. This suggests that the benefits gained from being helped by others are greater than the losses of helping others in terms of survival of an individual. If this is the case, then individuals who exhibited traits of altruistic behavior may have had an advantage in survival and reproductive success when compared to ones who did not exhibit those traits, leading to those genes being passed on. This theory, however, does not account for altruistic behavior among complete strangers who are not likely to be seen again after they instance of helping. If direct reciprocity was really the mechanism for how altruism was selected for, then one would only expect to see helping behavior among individuals who were likely to meet again.

A different theory is that of kin selection which suggests that while altruistic behavior may cause lower reproductive fitness for the individual engaging in the behavior, the benefit it would bring to *related* individuals who were helped would be large. In this case, altruistic behavior would benefit the passing down of genes from *related* individuals, causing for their traits to indirectly be passed down. This theory, however, also limits altruistic behavior to related individuals, which runs counter to examples seen in real life such as helping friends or strangers. While people are more likely to help people they know (and less likely to help strangers, a concept known as social discounting), altruistic behavior is not solely limited to related individuals. Therefore, if kin selection is the mechanism through which altruistic behavior has evolved, helping of strangers may be an unintended consequence stemming from the way determine kin are determined (e.g., people that spend a lot of time together).

The third and perhaps most probable theory of how altruistic behavior evolved is called inclusive fitness. The evolutionary biologists supporting the theory of inclusive fitness propose that altruistic behavior evolved because exhibiting altruistic behavior makes individuals more attractive as partners and mates. Through a process of self-sorting, altruistic individuals are paired together and people who do not exhibit altruistic traits are paired together. Then, due to the survival benefits of having helpful individuals help each other when in need, and thus being viewed as more attractive by others, people with altruistic traits would have increased reproductive fitness. This would support the data that suggests people are more likely to help people they know, but also explain helping of complete strangers, as doing so would make individuals more attractive to others around them.

**Studying Altruistic Behavior in the Lab**

Researchers have developed two main experimental paradigms to study altruistic behavior in a lab setting and determine which of the evolutionary theories is most accurate. The two tasks are called the Dictator Game and the Ultimatum Game (Wu et al., 2019). In the Dictator Game, participants are given a hypothetical sum (often tied to the real compensation of the study) of money and asked to make decisions on how to split it between themselves and another person. Altruistic behavior is measured according to how much people give to the other person.

The Ultimatum Game adds a component of response to the dictator game. A hypothetical sum of money is given to person 1 and they must choose how to split it up between themselves and person 2. Person 2 has the choice of either accepting that offer and splitting the money as chosen by person 1, or rejecting it, leaving both with either a set sum or nothing, depending on the experiment. Altruistic behavior can then be studied by observing how much person 1 is willing to give person 2 or by observing how often person 2 accepts an offer that benefits person 1 more and person 2 less, when compared to the base offer.

**Neural Correlates and Proposed Mechanisms of Altruistic Behavior**

Utilizing neuroscience tools such as fMRI while participants are playing the Dictator or Ultimatum games has given researchers a look into what neural processes are active while making altruistic decisions. Neuroimaging experiments highlight elevated BOLD responses in three main regions: the ventromedial prefrontal cortex (vmPFC), the ventral striatum and the right tempo-parietal junction (rTPJ), during while participants are making decisions about how altruistic to be (Hutcherson et al., 2015).

The vmPFC and ventral striatum are a part of the dopaminergic system in the brain, which is associated with value through predicted rewards, and reward learning (Haber, 2011). These areas also show elevated levels of BOLD activity during decision making experiments (Haber, 2011). Additionally, the vmPFC is associated with optimal action selection, with patients who have lesions in the area struggling to choose which action would be most beneficial when faced with multiple option for how to act (Bechara et al., 2000). This suggests that the vmPFC and ventral striatum are responsible for the value computation and action selection part of altruism, but the question remains of how this value computation is done. Clearly altruism is linked to decision making processes. However, the traditional cost/benefit weighting of evidence seems inadequate to explain altruistic behavior, suggesting an additional component in the value calculation. Furthermore, researchers have found that decision making times for altruistic decisions are longer than those for non-altruistic ones, supporting the idea of them being more computationally demanding (Hutcherson et al., 2015).

Some researchers suggest this additional component involved in value calculations relating to altruism is linked to empathy (Krebs, 1975). This would make sense as the rTPJ is linked to a network in the brain which is active during tasks requiring participants to imagine how they would feel if they were a person described by the researcher (DiNicola et al., 2020). However, this network is unique to humans and therefore, would not explain the cases of altruistic behavior observed in non-human animal species, such as rats.

**Original Model**

In an effort to shed more light on what the additional computational component in altruistic decision making could be, Hutcherson et al (2015) modelled altruistic decision making using a drift diffusion model (DDM) and compared its behavior to human behavior and brain activity patterns. DDMs are a type of model which compute a random decision variable over time using evidence and noise signals. DDMs are well sorted for describing altruistic behavior as they capture the stochastic nature of evidence accumulation from past experiences as well as sensory information. Furthermore, DDMs are model free decision-making models which means that they do not use explicit simulations to guide decision making. Model free decision-making has also been shown to be used in altruistic decision-making (Lockwood et al., 2020). Therefore, DDMs are relevant for studying altruistic behavior. The time component of DDMs allows for researchers to make response time predictions, which has been an area of interest for studies of altruistic behavior.

The researchers had participants play a modified version of the dictator game where they randomly presented the Active Participant (AP) with a predetermined monetary offer for themselves and a Passive Participant (PP) and were asked to make choices between accepting the presented offer or taking a baseline half and half offer. The researchers also ran model simulations for half of the trials for each subject. The decisions made by the AP are similar to what person 2 would have experienced during the Ultimatum Game, although there was no component of thinking about how the other chose to split the money (since offers were predetermined), meaning that the speculative motives of the proposer do not play a role in participants decisions.

The researchers collected data from human participants who each made 180 money allocation decisions while in an fMRI machine. Participants were randomly presented with one of nine combinations for proposed ways of splitting up the money between themselves and another person (for a total of 20 times per offer). The researchers added noise of to the proposed values to limit habituation of participants choices. Participants responded to each prompt on a 4-point Likert type scale with 1 representing a strong “no” and 4 representing a strong “yes” to the proposal. Each decision was followed by chance-based implementation of the decision wherein there was a 60% chance of the decision being implemented and a 40% chance the opposite decision being implemented to account for mistakes in decision making. Participants were told to ignore this and respond with the option they most preferred as that would have the highest chance of being implemented. At the end of the experiment participants were awarded the outcome of one randomly selected trial.

The researchers found that participants made altruistic choices on average of 21% 18 trials losing an average of $3.73 4.64 per trial and that the decision making time was higher on those trials where participants made altruistic choice with the average decision making time being 2300 ms 310 for altruistic choices and 2131 ms 280 for selfish choices (paired t(43) = 4.97, p < 0.0001) (Hutcherson et al., 2015). They also found that the BOLD responses in the vmPFC and TPJ were correlated with differences in altruistic behavior, with higher levels of altruistic behavior being linked to an increased BOLD response.

A modified DDM was developed by the researchers to try and model this altruistic behavior. The value calculation at each time point was calculated using difference equation.

where and are the amounts given to self and another and and are constant weight parameters. is gaussian noise that it added to the signal. The DDM had its threshold boundaries determined through an equation allowing for collapsing or static bounds. where b is the initial height of the barrier, and d is a decay parameter allowing for collapsing bounds to be implemented. A constant non-decision time (NDT) time period parameter was added to the beginning of each trial to account for the time it takes for human participants to begin making a decision.

The DDM had its parameters tuned by using part of the human participant data and a maximum likelihood approach. The model was then run to see how well it fit out of sample data. The researchers found that the model fit out sample data donation with a correlation coefficient of r = 0.94, reaction time with r = 0.96 between individuals. The model also did fairly well in predicting differences within individuals with a mean r = 0.88 for acceptance rates and mean r = 0.53 in response times.

Hutcherson et al., (2015) discusses how the model ties to neural structures in the brain. They found that the values $Self were correlated with BOLD responses in regions including the vmPFC and ventral striatum. On the other hand, values $Other were correlated with activity in the rTPJ. This suggests that the value calculations for values of self and other are represented independently in the brain (Hutcherson et al., 2015).

**Limitations of the Original Model**

A limitation of both the study Hutcherson et al., (2015) and the paradigms used to study altruism in the lab, in general, is that often they limit themselves to altruism in situations of gain. While this is relevant to explain, for example, the behavior of a monkey who lets a strange individual out of a cage to share food, it does not explain altruism in the context of loss. Many altruistic choices happen in the context of losses. For example, in situations such as saving a drowning child there is no tangible reward that either party gets from the interaction, rather the child being saved endures less of a loss, while the person saving them takes a significant risk while doing so. Similarly, for example, deciding to pay for someone’s groceries at the store when you notice they have left their wallet at home is an example of an altruistic choice for which there is no reward for the person making it.

It has been shown that humans are disproportionately affected by “negative” events relative to some positive baseline (Feng et al., 2021). This demonstrates a negativity bias, or rather that humans consider negative effects more than positive ones, a concept supported by the evolutionary function of avoiding harmful or potentially detrimental situations (Feng et al., 2021). This concept is especially relevant when considering human decision-making. As a key component of Prospect Theory, loss aversion indicates that when making decisions, particularly in contexts with uncertainty, humans tend to be more cautious of losses relative to their equivalent gains (Sokol-Hessner & Rutledge, 2018; Kahneman & Tversky, 1979). Thus, many studies have investigated altruistic behavior in terms of framing choices as losses vs. gains (e.g., Fiedler & Hillenbrand, 2020; Lusk & Hudson, 2010; Neumann & Böckenholt, 2014). Lusk and Hudson (2010), suggest that conducting the Dictator or Ultimatum games in the context of loss, shows a significant decrease in participants’ levels of altruistic behavior. The experiments conducted in this paper are thus based on a set of literature surrounding the loss/gain domains, previous experiments in neural economics, and studies on human prosocial behavior in the context of decision making. Specifically, we compare the various results observed in the loss and gain domains to see how altruistic behavior differs in these contexts.

It has also been shown that losses are represented differently in the brain than gains (Häusler, 2016). This is the case for both reward prediction and prediction error responses, suggesting that the computation for loss may happen via different neural circuits than that for gain. Furthermore, losses may be more emotionally charged than gains, activating limbic networks which could further support the idea that losses are computed differently than gains in the brain (Martino, 2010).

**Extensions to the Original Model**

The subsequently described experiments focus on expanding three main parts of the existing literature. Firstly, we aim to investigate how well the model by Hutcherson et al (2015) models human behavior in the domain of loss, by adapting it and applying it to a new dataset containing values framed as losses instead of gains. In this way, we plan to predict behavioral patterns of altruistic behavior in the loss domain and to see how well the model fits to human behavioral data. As losses are represented differently than gains in the brain, it will be interesting to see whether the evidence weighting calculation is able to model trends shown by human subjects, specifically those suggesting people are less altruistic in situations involving losses instead of gains, as well as tend to act faster for decisions involving losses.

Second, we plan to adapt the model to include a social distance factor to better describe the difference in observed altruistic behavior between familiar and unfamiliar individuals. The existence of some type of social distance factor in the evidence weighting mechanism would make sense if the main evolutionary mechanisms for altruistic behavior favored acting altruistically towards people that they are likely to receive help back from, as suggested by direct reciprocity, and kin selection theories. Modelling this as a parameter in the evidence weighting calculation is interesting as it could shed light on whether familiarity affects the evidence weighting or the threshold for making a decision.

Finally, we plan to extend both factors to the context of decision making under uncertainty to better simulate the uncertainty in gain and loss in the context of altruistic choices in the real world. The observed differences of in outcomes will be interesting to observe as most real-life altruistic decisions contain a large portion of uncertainty of what the outcome of that decision may be (e.g., uncertainty of possible danger in situations of protecting others).

We hypothesized how we expected our models to work prior to modifying the model developed by Hutcherson et al. (2015). In the contexts of losses, we expected the model to favor rejecting the altruistic offers based on human data as well as the how the losses are represented in the equation. For the social distance factor, we expected the acceptance likelihood in the gain domain to increase as the social distance factor grew (as people were more familiar) and the acceptance rate to decrease in the loss domain (as people wanted to avoid losses). For decisions under uncertainty, we expected response times to increase and acceptance rates to decrease with increased risk, as well as decrease with increased social distance.

**Methods and Results**

**Changes to the Evidence Weighting Calculation**

The original DDM model proposed by Hutcherson et. al utilizes the following difference equation representing the stochastic relative decision value signal that accumulates over time .

We expand this model so that it can both be generalized to different data sets (namely, contexts of loss) and can incorporate additional parameters of interest, namely uncertainty and social distance. The motivation behind expanding this model to the loss domain is dependent on behavioral evidence that shows a particular form of loss-calibration known as “utility-weighted inference” accounts for somewhat irrational biases involved in human decision-making processes. We aim to reflect this loss-averse behavior using our model to account for differences in preferences across these domains (Lieder et al., 2018). Note that other studies in neuroeconomics incorporate loss-level utility functions, but for the scope of our experiments we restrict our studies to loss-framed datasets (e.g., Boqvist et al., 2011; Zhao et al., 2020).

Our new model also incorporates additional key variables, , , , representing the social distance factor, the probability of a proposed offer for the Self, and the probability of a proposed offer for the Other. We choose to utilize the DDM as a baseline to account for the observational correlations between temporal discounting and social discounting found by Osiński et al., (2015) and expand these observations to the social distancing sphere as the DDM allows for response time considerations. We focus our studies on social *distance,* rather than looking at social *discounting* factor for the following reasons. Social *discounting* is a phenomenon in which people are willing to sacrifice more resources for those that are socially close to them rather than distant others. However, in the context of altruistic behavior, we look at the impact of social (or relational) *distance*, as researchers propose that increased distance between the Self and Other allows for reduction of self-serving considerations and frees the decision-making process of the egoistic Self (Osiński et al., 2015). Thus, factoring in a social *distance* component can allow us to better isolate and understand how people make decisions for others in an altruistic setting. This is also motivated by behavioral studies that have shown that as social distance increases, risk-seeking behaviors increase as well (Guo et al., 2019). Furthermore, current literature indicates that risk-seeking behaviors and prosocial behaviors differ in loss and gain domains (Feng et al., 2021). Thus, we incorporate a social distancing factor to observe results in our novel contexts, particularly the intersection of loss/gain domains, social familiarity, and altruistic behavior. Additionally, human subjects self-report relational distances between themselves and others as numerical quantities. Thus, it seems fitting to represent social distance as an integer quantity in accordance with this behavior (Zhang et al., 2017).

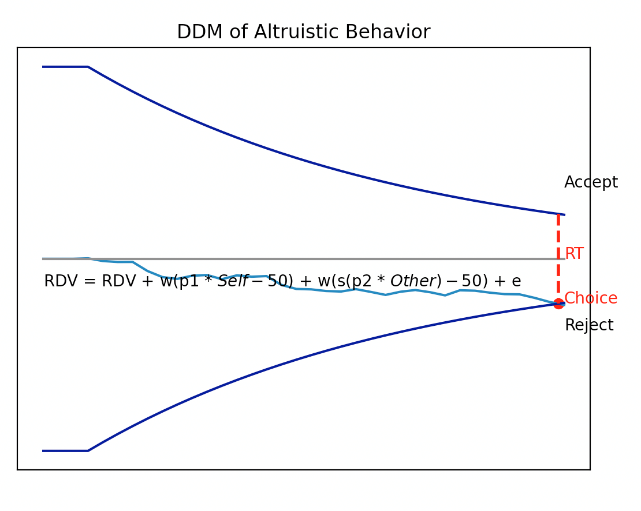
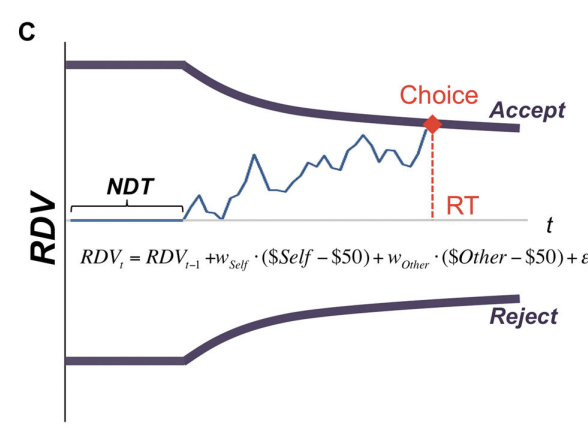
Internal representations of uncertainty have been considered in one of two ways, either at the level of decisions or in a Bayesian manner across the perceptual pathway (Koblinger et al., 2021). A large class of models has been used to study uncertainty in decision-making at all levels of the process. Here, we use a non-probabilistic recognition model as a type of complex generative model. We allow for the probabilistic computation of the observational variable in a task-dependent manner. Thus, when an offer is presented with a certain value, the value of the information is weighed relative to its probability, giving larger values with higher probabilities overall higher relative values, while large values with lower probabilities will yield lower results. This is manifested in our model by the incorporation of probabilities as a multiplicative factor when computing relative decision values for presented offers. Thus, larger offers with greater uncertainty are viewed differently than guaranteed offers and are represented accordingly in the decision-making process.

We also generalize our base offer using the parameters and so that we can apply our model to different data sets that don’t necessarily propose a sure loss or gain of $50. Thus, in the subsequent experiments, we use the difference equation below with parameters identical to those used by Hutcherson et. al save for values of , , , , and .

For the purposes of our experiments, we used the mean best fitting DDM parameters for each subject found in the Hutcherson et. al. study. For each experiment, we simulated subject-level consideration of each of the possible offers proposed, running 1000 trials at each offer, keeping track of average response times (time until a decision is made) and generosity. While typical studies utilizing DDM models contain real-life subject participation and then use that information to get the best-fitting DDM parameters, we did not have access to this kind of information for our experiments. Thus, we instead ran our model for 1000 trials for each offer presented to avoid variability in results and used the mean best-fit parameters from the original study for the constant (not free) parameters. For the following discussion, we assume that individuals generally have a preference for self, setting and . Note that for “altruistic” individuals, best-fitting DDM parameters could yield varying weights, indicating greater preference for the other, but for the analysis below, we assume these values as controls. Additionally, we set the NDT for the remainder of our experiments to be 868 although this factor has no effect on the qualitative results observed. As acknowledged in the original study, the average NDT is larger than the usual for DDMs due to the additional processing time required to determine the payoffs from each trial. However, since our data sets involve the computation of payoffs for each offer presented, we use this larger NDT time to account for the sensory and motor-related processes that must be factored into the pre-computation decision making time. We also use standard values for both b, the starting decision threshold value, and d, the collapse rate of the decision threshold, and respectively.

**Figure 1**

*Comparison of single trial behavior for model by Hutcheson et al. (left) and our replication (right)*



**Replication of the Original Model**

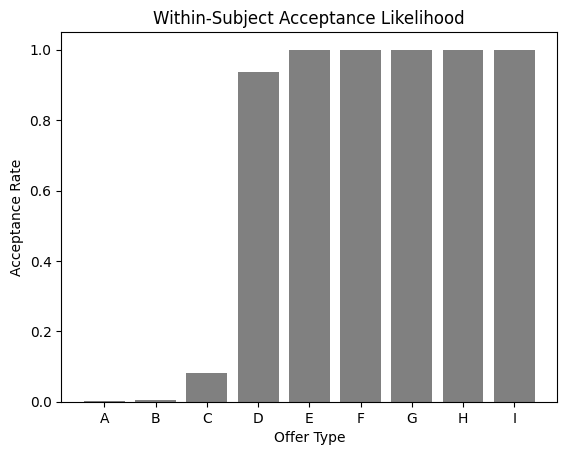
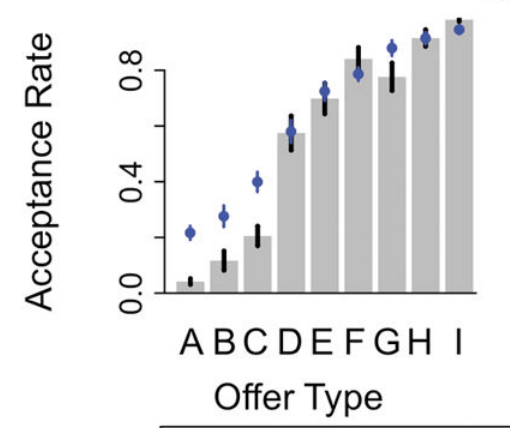
As seen from figure 1, when we visualize the decision-making process for one offer, both the Hutcherson et. al. model and our new proposal, shows similar structure in evidence weighting within the multi-attribute drift diffusion model.

As a sanity check, we can set to values of 1, simulating either a proposed sure gain and set to since we want to subtract the proposed offer from the base. We also set the social distance factor, to 1 so that only the weights for Self and Other affect the decision. A comparison of results from our new model and the results from the Hutcherson et. al. study indicates that our model still accounts for the results observed using the difference equation proposed in their paper but allows for additional explorations. Note that as in the original paper, we use the following offer proposals described in the legend for figure 2.

Both models show the same inclination towards offers that benefit the self, although we still see some altruistic behavior (see figure 2 for average acceptance trends). Note that variability in results likely stems from the difference in weights, collapse rates, and threshold values used, but the overall generalizations that can be made apply in both cases. Similarly, we see the same increase in response times for generous choices, both across offers and when compared to selfish choices (see figure 3 for average response time trends).

**Figure 2**

*Comparison of within subject acceptance likelihoods over multiple responses for the model and behavioral data by Hutcherson et al., (model run for 90 trials after selecting best-fit parameters from behavioral data) (left) and our replication after 1000 trials (right)*



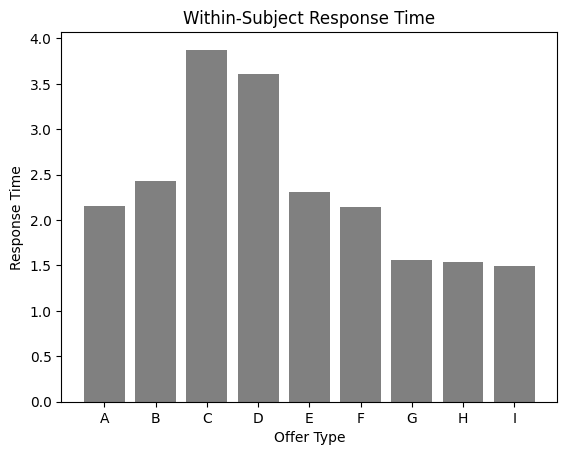
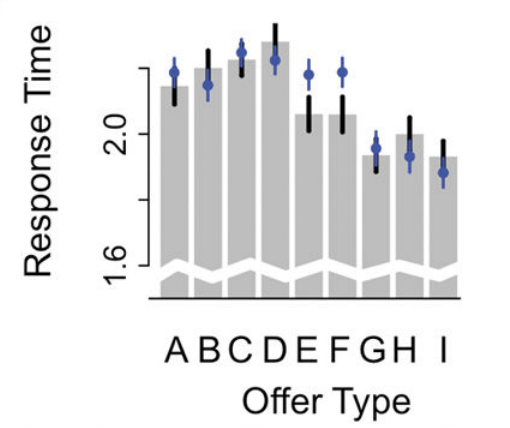
A: ($25, $75)     B: ($25, $100)   C: ($50, $10)

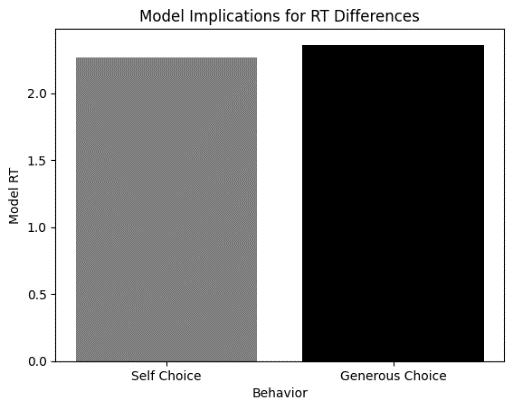
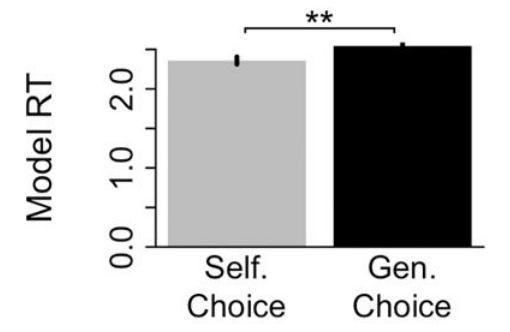
Legend: D: ($50, $100)  E: ($75, $10)     F: ($75, $25)

G:($100, $10)   H: ($100, $25)   I: ($100, $50)

**Figure 3**

*Within subject average response time per decision prompt for model and behavioral data by Hutcherson et al and our replication after 1000 trials*





**Incorporating Loss Contexts**

In order to expand our model to the loss context, we applied this model to a new dataset in which each of the offer types was a proposed loss rather than the gain and was compared to a baseline proposal of losing . Since relative value is calculated in comparison to the baseline proposal, we use so that loss amounts less than the baseline yield positive results, making decisions in favor of those offers more likely.  Thus, offers that were previously considered “generous” (providing the possibility of giving the Other more money and having less for the Self)  now became “selfish” since a proposed loss of for Self, relative to baseline, is better than a proposed loss of for Other relative to baseline, for example (

**Table 1**

*Values presented in each case for loss scenarios*

A: ($-25, $-75)     B: ($-25, $-100)   C: ($-50, $-10)

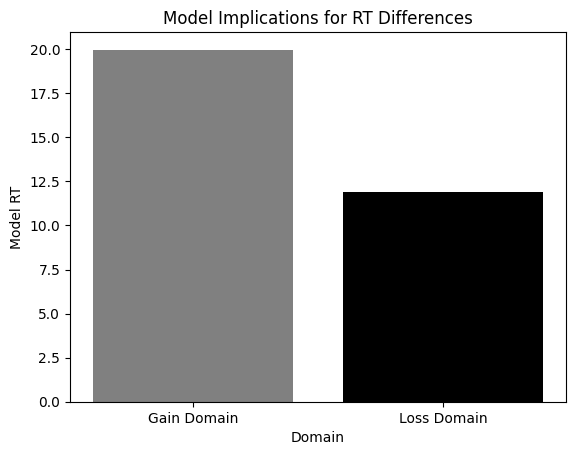
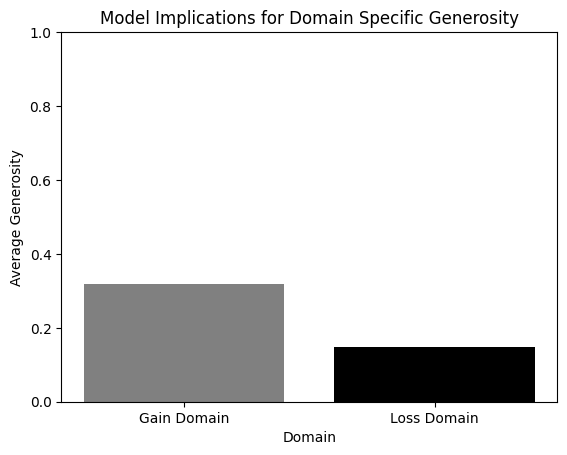
D: ($-50, $-100)  E: ($-75, $-10)     F: ($-75, $-25)

G:($-100, $-10)   H: ($-100, $-25)   I: ($-100, $-50)

We define generosity in terms of cases where a proposed alternative that was not better than the baseline offer for the Self (but is better than the baseline offer for the Other) is chosen. Average generosity is computed over all trials in which a generous choice is chosen over all trials in which a generous alternative is presented. We find that the overall results have many implications for understanding altruistic behavior in both the loss and gain domains. Notably, when considering overall generosity, we find that in the loss domain, subjects are less likely to demonstrate altruistic behavior relative to altruistic behavior observed in the gain domain ( vs. . These findings are consistent with prior research indicating that losses loom larger than gains, and that individuals may demonstrate greater loss aversion in loss contexts, and thus show less altruistic behavior.

**Figure 4**

*Differences in generosity rate and response time for gain and loss domains of altruistic choice behavior*



One possible framework for subject behavior in the loss/gain domains includes a proposed 2 system-framework of information processing in the brain in which System 1, the intuitive system, consists of decisions made based on fast and automatic processing whereas System 2, the deliberative system, spends greater time on slower, more rational processing (Kahneman, 2011). As such, using the response times obtained from the DDM model allow us to consider these systematic processing methods in the context of decisions made in both the loss and gain domains. We see that subjects demonstrate less altruistic (or generous) behavior in the loss domain, as mentioned above, in addition to showing lower response times for decision making This implies that perhaps decision-making in the loss domain may be affective and automatic, yielding short-term, egocentric decisions that are influenced by fears of loss, rather than the more deliberative response times demonstrated in the gain context. We also note the differences in response times between generous and selfish choices as presented by Hutcherson et. al. may be caused by these different processing mechanisms, as generous responses typically should take longer (and do, in model simulations).

**Incorporating Uncertainty in Outcomes**

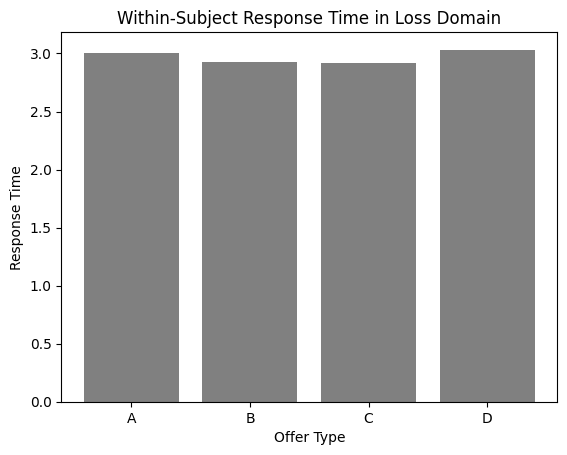
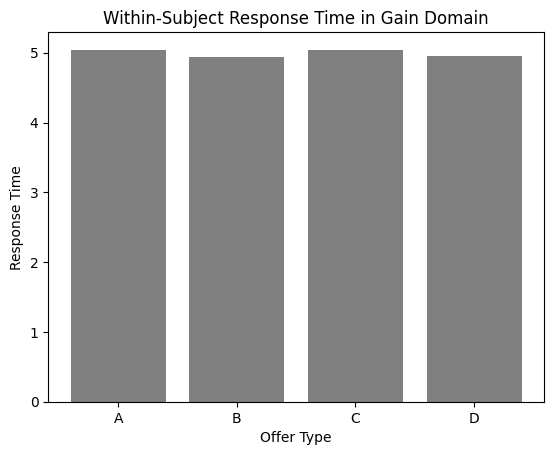
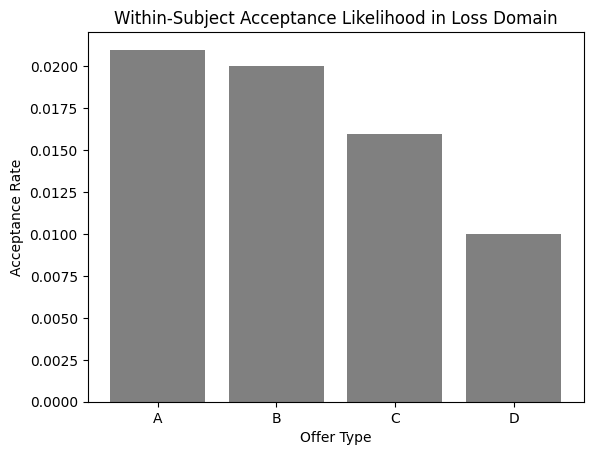
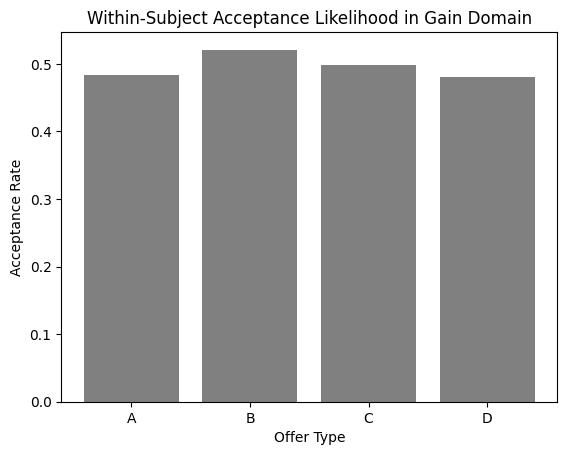
Paradigm studies of loss aversion typically focus on human decision-making between sure losses and potential losses. For example, in a mixed offer trial, people typically don't choose gambles with an equal probably of gaining $30 compared to losing $20 unless the amount of gain is roughly twice the amount of potential loss (Zhang et. al. 2017), indicating that losses are weighed roughly twice as strongly as gains. We expand our exploration of loss aversion to also account for framing effects on risk preference in accordance with classical studies in this domain. Thus, the next experiments aim to address questions of risk preference in loss and gain domains. Namely, we consider how loss aversion affects decision making under situations dealing with uncertainty, since individuals are often faced with choices that are more probabilistic in nature.

We simulate the experiment presented by Zhang et. al. in 2017 in which human subjects were presented with a modified version of the cups task consisting of gain and loss domains. Subjects are given a choice between a sure gain or loss of $5 and a designated probability of a larger gain or loss. In the experimental setting, this was implemented by asking subjects to choose between two sides of a screen, one which contained a cup that guaranteed a certain amount of money being gained or lost, and another which contained multiple cups with only one of the cups containing the specified loss or gain. Note that the expected value of return in each case is the same for both risky and sure options, indicating that any results seen reflect subject preferences (Zhang et. al 2017).

The incorporation of probabilistic factors for alternatives ( in our model) allows us to simulate this experiment using computational tactics. We assume our base value proposal is a guaranteed loss or gain, and the proposal is considered using some factor of uncertainty. As in the previous set of experiments, we run 1000 trials in which each offer is presented, keeping track of average response times, acceptance rates, and generosity. Save for the probabilistic components, offer values, and base values, the parameters are identical to the previous set of experiments. Note that we also use the same offers for both and to replicate the Zhang et. al. experiment as closely as possible. Thus, we utilize the following offer combinations (with the signs flipped in the loss domain trials) (see figure 4 for results).

**Figure 4**

*Acceptance rate of offers for experiments including probabilistic component*



Once again, we find that loss domain trials take significantly less time than gain domain trials, with an average response time for the gain domain ranging from to and to . This again indicates that it may be likely for automatic processing in which negative emotions towards potential losses account for why loss domain trials take less time. Notably, however, we find some differences between our results and the results observed in the study. Rather than seeing participants being more risk averse in the gain frame, we find that participants are risk seeking towards potential gains but are less risk seeking when it comes to potential losses ( This implicates the social distance component considered in the original Zhang et. al. study. Rather than using only a pair in which the active participant (or dictator) is making decisions for a complete stranger, the study explores how making decisions for different individuals with different social proximities affects both loss aversion. Other studies exploring the impacts of social distance on decisions for Self and Other inform the following extensions (Feng et al., 2021; Osiński et al., 2015; Zhang et al., 2017) We expand this to account for altruistic behavior as well within the loss domain.

**Accounting for Social Distance**

We incorporate a multiplicative social distancing factor that considers the option presented to the $Other with greater weight. Note that in our original studies, we utilized a baseline of weights for and in which the self was in general weighed more heavily than the other. Thus, the incorporation of social distance as a multiplicative factor can be thought of an an extension of the original weights being placed upon decisions made for the self and for the other, although we choose to keep this factor as a separate component rather than to be factored into the calculator for the weights to better isolate the effects of social distance on altruism and decision making in the loss and gain domains. The following experiments still use the baseline values of and as weights for the Self and Other, although we pick a social distance of as our standard for comparing generosity across trials. Definition of generosity is the same as above and uses the offers presented in the Hutcherson et. al. (2017) study as there is a clearer way to define generosity, and therefore study altruism, with those offers.

We use the range of social distances presented by Osiński et al., (2015), namely considering positions lower than 10 “socially closer” or closest to self. We find that when we incorporate a social distance factor of 5, overall generosity increases in both the loss and gain domains (see figure 5 for results). Namely, more generous offers are accepted almost in all cases when the social distance component is incorporated for *close others*. These findings allow us to simulate human behavior in the context of extremely altruistic actions for close others. Namely, as presented above, donating a kidney to a family member, mothers selflessly saving their children, and more (Stromberg, 2014). While we focused our model to account for these behaviors in order to demonstrate the impacts of social distance on altruistic behaviors in both loss and gain contexts, additional components can be incorporated here to further study the effects of social discounting in addition to social distancing, and the difference of subject-level generosity between altruistic individuals and others depending on the starting weights for Self and Other.

Once the social distancing component was tested, we conducted trials in the context of offers presented by Zhang et. al. We ran 1000 trials at each of the offers presented for each of the social distancing factors ranging from 1 to 100. Note that we can think of social distance as proximity to subject, with some research using subject interpretation of physical distances to measure proximity, or self-reported closeness. Since most existing research uses an integer range of 0 to 100 to account for social distance, here we also simulate using these values. We propose that a social distance of 1 is akin to being viewed as Self, a social distance of 5 implies a close relative, a social distance of 10 a close friend, and so on. Since we are interested in kin selection, we used a social distance of 5 for the previous studies, but we consider all possibilities below to fully understand the impact of social distance on acceptance likelihood of risky prospects in both the loss and gain domains.

**Figure 5**

*Model generosity rates in gain and loss domains of altruistic choice*

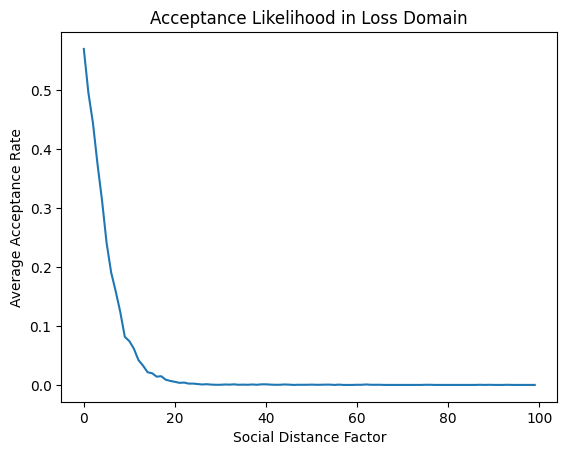
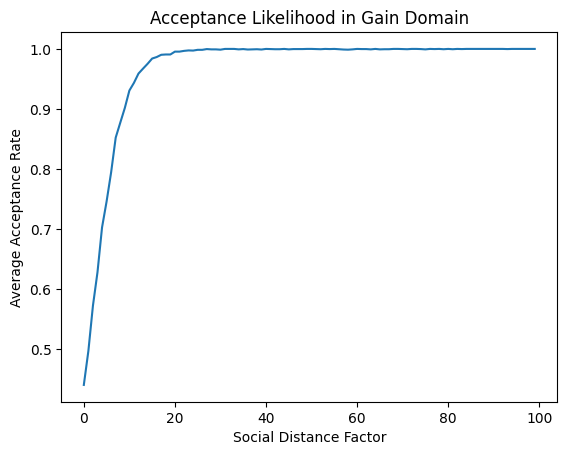
Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

**Figure 6**

*Model offer acceptance rate as a function of social distance in gain and loss domains of altruistic choice*



As found in the Zhang et. al. study, we have two key findings: subjects are more risk averse in gain situations when making decisions for close targets rather than distant targets and are more risk seeking in the loss situation when making decisions for proximal targets compared to distal targets (Zhang et. al. 2017). Namely, we see that as social distance increases, loss aversion increases as well. While this may seem unintuitive with respect to social discounting, the model indicates that in the loss domain, risky choices are less likely to be taken for complete strangers rather than for self. Since human behavior demonstrates that a sure loss is often much less preferred than a probabilistic loss, we see that decisions made for Self and Other when the social distance factor is low, are close to decisions made for self. Thus, for ourselves, and those close to us, we tend to demonstrate greater loss aversion. Computationally, this makes sense since lower social distances imply that offers for Self and Other are computed the same, save for differences in weights. Behaviorally, this also makes sense since loss aversion mainly applies when considering utility for Self. Thus, a small guaranteed loss for a complete stranger is less appealing than a guaranteed loss for Self. Possible explanations for this behavior may be that people don't want to be blamed if complete strangers face a larger loss as a result of their actions or that the sure small loss seems “less risky” than a probable greater loss when considering for others, whereas loss aversion for the self makes it seem “more risky” even if expected values are the same.

**Discussion**

**Conclusions and Implications**

By adapting the model created by Hutcherson et al. (2015) we were able to extend its applicability to altruistic behavior in the contexts of loss, familiarity and uncertainty. As hypothesized and suggested by behavioral data from Hudson et al., (2010) we found that our model had lower generosity rates in the contexts of loss. This suggests that the decision variable computation is biased in away that favors not acting altruistically when events are seen as leading to losses. Additionally, our model supported idea of altruistic decisions being instinctive and automatic rather than reasoned (Rand & Epstein, 2014), by showcasing that average reaction times for decisions in the context of loss were lower than those in the context of gain.

In the field of familiarity, we were able to successfully implement a parameter, which showcased that people acted more generously towards people they were more familiar with as suggested by evolutionary theories supporting our hypotheses. Interestingly, our model showcased that in the loss domain acceptance rates in the were similar in cases for both familiar and unfamiliar individuals, suggesting that social distance has less of an effect when choices are framed as losses.

In decisions containing probabilistic components we saw similar trends in that acceptance rates were higher for gain contexts and response times were lower for loss contexts as hypothesized. When incorporating probability of different outcomes into decisions we saw overall acceptance rates decrease and response times increase.

When combining the probabilistic component and social distance, our data supported the idea that people were less likely to make decisions involving risk in the gain domain when social distance was low, meaning the other was more familiar. This could suggest that if there is already a guaranteed smaller gain people are likely to want to take it when with someone they know, as they are averse to taking responsibility for possible loss for people they know. In the loss domain our data showed that people were more likely to take the risks of losing less when dealing with people they knew. This could suggest support an idea that people are more likely to want to take a chance to save others from risk when they know the other.

**Limitations and Future Extensions**

Looking forward, while this model accounts for many of the discrepancies between choices made for Self and Other in both the loss and gain domains, particularly in the context of altruistic behavior, there is much research to be done to account for behavioral differences across subjects. For starters, our model assumed a baseline of non-altruistic subjects, namely using higher weights for Self as compared to Other to account for the greater value placed on self-interest. While this does not diminish altruistic behavior, utilizing this baseline fails to address altruistic differences that may be present across subjects, and specifically across cultures. It may be of interest to compute weights based on some consideration of socio-cultural-religious values to observe the effects on prosocial behaviors in these contexts, as research has shown that cultural differences make significant differences in generosity (Yablo & Field, 2007).  Namely, possible future experiments could model the data generated from participant administration of Self-Report Altruism (SRA) scales finding that in Eastern cultures, specifically Thai individuals, scored significantly higher than those from the U.S. (p<0.001). Thus, running a fitting a DDM to get best-fitting parameters on subjects across cultures and calculating the mean parameters may indicate differences in baseline altruism depending on socio-cultural factors, allowing for additional explanations of altruistic behavior across individuals.

Additionally, while our model explored the social distancing component in an attempt to account for generosity towards complete strangers, social discounting has also been observed in experiments with human subjects. While we focus on the familiarity components, reflecting an integer-level representation of human relations, it is possible to also frame this as a social discounting factor, meaning that not only will the Self be valued more, but also that as the social distance increases, the consideration of others’ welfare decreases. This social discounting could be interpreted as a parameter in which considerations for the self are less than a factor of 1, and a best-fitting DDM for subject-level experiments could test whether the reliability of this kind of component exists. Ultimately, due to the way that task-dependent computations in decision making are represented independently in the brain, there are many components that can be isolated and studied to better understand human behavior and its internal representations. While we provide some possible extensions above, much of the literature in this domain is novel and there exists many areas where our understanding can be furthered.

We represent social distance as an integer from 1-100 that is factored into the computation when considering effects for the Self and Other. While these results reflect the differences in risk-taking behaviors observed in both gain and loss domains ranging from close others to complete strangers, using a factor between 0 and 1 in accordance with social discounting (as explained above), and then possibly modifying this value based on the social distance (represented by us as akin to the physical Euclidian distance between a simulated representation of two subjects as a representation of relational distance and separate from the discounting factor), may be more accurate for simulating generosity, as we generally do not see as significant levels of generosity in humans within the gain domain for complete strangers unless an explicit reciprocity or other incentive is apparent. Thus, additional manipulations of parameter values may yield results that better fit with studies in the sphere of altruism, although for the sake of our discussions, since we were studying close social distances, the qualitative results still reflect human behavior of altruistic choices for close others (kin-selection).

Further studies could also use human or animal subjects to find best-fit parameters for altruistic behavior in the context of losses making for more direct comparison between behavioral and computational data. This would allow for further tweaking as well as validation our model using accurate scaling.

**References**

American Psychological Association. (2015). *Altruism May Be Universally and Uniquely Human*. https://www.apa.org. https://www.apa.org/pubs/highlights/peeps/issue-48

Bechara, A., Tranel, D., & Damasio, H. (2000). Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain*, *123*(11), 2189–2202. https://doi.org/10.1093/brain/123.11.2189

de Waal, F. B. M., Leimgruber, K., & Greenberg, A. R. (2008). Giving is self-rewarding for monkeys. *Proceedings of the National Academy of Sciences*, *105*(36), 13685–13689. https://doi.org/10.1073/pnas.0807060105

DiNicola, L. M., Braga, R. M., & Buckner, R. L. (2020). Parallel distributed networks dissociate episodic and social functions within the individual. *Journal of Neurophysiology*, *123*(3), 1144–1179. https://doi.org/10.1152/jn.00529.2019

De Martino, B., Camerer, C. F., & Adolphs, R. (2010). Amygdala damage eliminates monetary

loss aversion. *Proceedings of the National Academy of Sciences*, 107(8), 3788–3792.

https://doi.org/10.1073/pnas.0910230107

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785–791.

https://doi.org/10.1038/nature02043

Guo, H., Song, H., Liu, Y., Xu, K., & Shen, H. (2019). Social distance modulates the process of

uncertain decision-making: evidence from event-related potentials. *Psychology Research*

*and Behavior Management*, *Volume 12*, 701–714. https://doi.org/10.2147/prbm.s210910

Haber, S. N. (2011). Neuroanatomy of Reward: A View from the Ventral Striatum. In J. A. Gottfried (Ed.), *Neurobiology of Sensation and Reward*. CRC Press/Taylor & Francis. http://www.ncbi.nlm.nih.gov/books/NBK92777/

Hutcherson, C. A., Bushong, B., & Rangel, A. (2015). A neurocomputational model of altruistic choice and its implications. *Neuron*, *87*(2), 451–462. https://doi.org/10.1016/j.neuron.2015.06.031

Häusler, A. N., Oroz Artigas, S., Trautner, P., & Weber, B. (2016). Gain- and Loss-Related Brain Activation Are Associated with Information Search Differences in Risky Gambles: An fMRI and Eye-Tracking Study. *ENeuro*, 3(5), ENEURO.0189-16.2016. https://doi.org/10.1523/ENEURO.0189-16.2016

Kay, T., Keller, L., & Lehmann, L. (2020). The evolution of altruism and the serial rediscovery of the role of relatedness. *Proceedings of the National Academy of Sciences*, *117*(46), 28894–28898. https://doi.org/10.1073/pnas.2013596117

Koblinger, Á., Fiser, J., & Lengyel, M. (2021). Representations of uncertainty: where art

thou? *Current Opinion in Behavioral Sciences*, *38*, 150–162.

https://doi.org/10.1016/j.cobeha.2021.03.009

Krebs, D. (1975). Empathy and altruism. *Journal of Personality and Social Psychology*, *32*(6), 1134–1146. https://doi.org/10.1037/0022-3514.32.6.1134

Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision

making reflects rational use of cognitive resources. *Psychological Review*, *125*(1), 1–32.

https://doi.org/10.1037/rev0000074

Lockwood, P. L., Klein-Flügge, M. C., Abdurahman, A., & Crockett, M. J. (2020). Model-free

decision making is prioritized when learning to avoid harming others. *Proceedings of the*

*National Academy of Sciences*, 117(44), 27719–27730. https://doi.org/10.1073/

pnas.2010890117

Decision-Making and Extreme Altruism. *PLOS ONE*, 9(10), e109687.

https://doi.org/10.1371/journal.pone.0109687

Roeder, J. (2013). *The Altruistic Pigeons – 10,000 Birds*. https://www.10000birds.com/the-altruistic-pigeons.htm

Sanders, M., & Tamma, F. (2015, March 23). The science behind why people give money to charity. *The Guardian*. https://www.theguardian.com/voluntary-sector-network/2015/mar/23/the-science-behind-why-people-give-money-to-charity

Sato, N., Tan, L., Tate, K., & Okada, M. (2015). Rats demonstrate helping behavior toward a soaked conspecific. *Animal Cognition*, *18*(5), 1039–1047. https://doi.org/10.1007/s10071-015-0872-2

Stromberg, J. (2014, October 15). *The science of extreme altruism: Why people risk their lives to save strangers*. Vox. https://www.vox.com/2014/10/15/6973971/extreme-altruism-risk-life

Wang, C., & Lu, X. (2018). Hamilton’s inclusive fitness maintains heritable altruism polymorphism through rb = c. *Proceedings of the National Academy of Sciences*, *115*(8), 1860–1864. https://doi.org/10.1073/pnas.1710215115

Wu, J., Balliet, D., Kou, Y., & Van Lange, P. A. M. (2019). Gossip in the Dictator and Ultimatum Games: Its Immediate and Downstream Consequences for Cooperation. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/article/10.3389/fpsyg.2019.00651

Yablo, p. D., & field, n. P. (2007). The role of culture in altruism: thailand and the united states.

Psychologia, 50(3), 236–251.  https://doi.org/10.2117/psysoc.2007.236

Zhang, X., Liu, Y., Chen, X., Shang, X., & Liu, Y. (2017). Decisions for Others Are Less

Risk-Averse in the Gain Frame and Less Risk-Seeking in the Loss Frame Than Decisions for the Self. *Frontiers in Psychology*, 8. https://doi.org/10.3389/

fpsyg.2017.01601

Zhao, W. J., Walasek, L., & Bhatia, S. (2020). Psychological mechanisms of loss aversion: A

drift-diffusion decomposition. *Cognitive Psychology*, *123*, 101331.

https://doi.org/10.1016/j.cogpsych.2020.101331

‌