

Human exploitation of social others' decreased relative fitness as a source of safety

Noah Guzmán

Computation and Neural Systems Program
California Institute of Technology

In this study, I investigate whether people exploit the weaknesses of their fellow humans as a source of safety when making decisions under threat. Building on Hamilton's selfish herd theory and other findings in the behavioral ecology of fear, I design a naturalistic and dynamic task environment founded on the notion of continuous decisions to disambiguate between multiple decision-making strategies using safety cues in the environment. Having deployed this task as an internet application, I find that the evidence does not support the hypothesis that humans willingly exploit the decreased relative fitness of social others in their safety decisions.

Introduction

There is a well-worn joke from the American novelist Jim Butcher about what to do in the event of a bear attack: "You don't have to run faster than the bear to get away. You just have to run faster than the guy next to you." Though just a joke, Butcher's statement reveals that human beings may be willing to exploit the weaknesses of others (in the case of the joke, a slower running speed) as a strategy to ensure their own safety.

The idea that we may view our fellow human beings as a source of safety is not a new one, and has been most thoroughly investigated in the literature on "social buffering". In the social buffering phenomenon, when an individual is faced with a threat, the perceived presence of another person or cues representing another person leads to a reduction in fear learning and threat-induced defensive reactions. For humans, the physical presence or perception of social support figures but not strangers prevents acquisition of fear associations^{1,2} When the presence of a social other during fear extinction is combined with exposure to a social other's safety behavior, recovery of learned fear is abolished.³ There is comparative evidence for similar phenomena across species; in one example in rodents, social company during fear extinction disrupts fear renewal.⁴ Social buffering has even been shown to reduce the intensity of pain and pain-related stress.⁵

The social modulation of threat responses and threat learning are more nuanced than the mere presence or absence of a social other attenuating fear learning. For example, the social buffering effect varies with the physical proximity and emotional bond between in-

¹ Hornstein, Fanselow, and Eisenberger 2016.

² Hornstein and Eisenberger 2017.

³ Pan, Olsson, and Golkar 2020.

⁴ Yuan et al. 2018.

⁵ Che et al. 2018.

dividuals.⁶ In one case, the presence of a caring relational partner led to more attenuated neural threat responses than the presence of a stranger.⁷ Additionally, higher perceived mutuality (the propensity of couples to view themselves as a dyad rather than distinct individuals) corresponds with decreased self-regulatory effort and attenuated preparatory motor activity in response to threat cues, even in the absence of direct physical contact with social resources.⁸

Most relevant for the present study is the idea that the competence of social others can signal safety to an individual. This has been demonstrated in humans: if a person believes their safety depends on the task performance of another, then perceiving that other person as competent at the task reduces anticipations of pain.⁹ This result suggests that inferences or perceptions of competence act as predictors of protection. In addition to reducing anxiety and the anticipation of pain, the perception of another's competence (a proxy for fitness) can bias decision making under threat.¹⁰

The hypothesis that competence is a predictor of protection and can act as a safety cue for conspecifics is further supported by the comparative phenomenon of anti-predator service in primates. In many species of primates, the males in a group tend to be more vigilant and skilled at detecting predators than females^{11,12,13} This competency at predator detection leads to reduced predation risk for the entire group, so in return for this service, males are granted inclusion into the group and all the benefits this entails (e.g. mating access to females)^{14,15} Thus, the females (and to a lesser degree, other males¹⁶) seek out competent males as a source of safety.

However, the opposite possibility suggested by Butcher's joke - that individuals might exploit the weaknesses of social others they perceive as incompetent for safety - has heretofore not been explored. In this study, I investigate whether and how human individuals might exploit a social others' decreased fitness relative to their own as a safety-seeking strategy. Theoretical work in behavioral ecology provides a foundation from which to develop explicit hypotheses regarding exploitative behavioral strategies as well as design experiments to test those hypotheses.

Hamilton's "selfish herd" theory^{17,18} and its empirical validations enable baseline predictions about how humans will behave when making decisions under threat in a social context. Selfish herd theory stipulates that when threatened by a predator, each individual in a group of organisms will move towards other individuals in the group so as to minimize their "domain of danger" (the neighborhood of unoccupied space around them)^{19,20,21} Assuming that a predator strikes randomly at the nearest prey, an individual adopting this strategy selfishly decreases their likelihood of predation.

⁶ Beckes and Coan 2011.

⁷ Coan, Schaefer, and Davidson 2006.

⁸ Coan, Kasle, et al. 2013.

⁹ Tedeschi et al. 2015.

¹⁰ Qi et al. 2018.

¹¹ Baldellou and Peter Henzi 1992.

¹² C. P. v. Schaik and Hörstermann 1994.

¹³ C. P. Schaik et al. 2022.

¹⁴ C. B. Stanford 2002.

¹⁵ Ribeiro da Cunha 2017.

¹⁶ C. Stanford 1998.

¹⁷ Hamilton 1971.

¹⁸ Eshel, Sansone, and Shaked 2011.

¹⁹ Morton et al. 1994.

²⁰ Morrell, Ruxton, and James 2011.

²¹ Viscido, Miller, and Wetthey 2002.

Though empirical data supporting the selfish herd theory have only been gathered on non-human animals^{22,23} the similarity between human defensive responses to threats and the defense behavior of non-human mammals²⁴ as well as overlap in the associated neural mechanisms^{25,26,27,28} supports the hypothesis that selfish herd theory also applies to humans.

When selfish herd theory is synthesized with the knowledge that the fitness of social others biases decision making under threat in a social context (and taking some inspiration from Butcher's bear attack joke), it is possible to formulate an ecologically realistic experimental paradigm for testing the hypothesis that individuals will exploit another person's lack of fitness (relative to their own) as a source of safety. The resulting experimental paradigm is a virtual ecology²⁹ combining a foraging task with a predator escape task that requires participants to make real-time *continuous* decisions regarding movement in the environment.³⁰ Using movement speed as a proxy for fitness, the speed of the predator and social other vary from trial to trial to be either slower or faster than the participant, with the speed conditions indicated to the participant on each trial. The task environment and experimental conditions are designed so that there are two potential strategies for successful performance of the task: stay close to the safety refuge or, following selfish herd theory, moving closer to the social other to reduce the domain of danger. A participant's chosen strategy on a given trial will likely be biased by the speed of both the predator and social other:

1. In the case where the social other is slower than the participant, the participant will choose the selfish herd strategy since they will still be able to outrun the social other to the refuge.
2. In the case where the social other is faster than the participant, the participant will choose to remain near the refuge since they cannot outrun the social other.

In the sections *Task Environment* and *Design*, the task dynamics, experimental design, and hypotheses will be covered more formally and in greater detail. For the moment, it suffices to say that current understanding of decision-making under threat in social contexts predicts a spectrum of decision strategies modulated by the relative fitness of both social others and the threat.

²² Viscido and Wetthey 2002.

²³ Orpwood et al. 2008.

²⁴ Caroline Blanchard et al. 2001.

²⁵ Christianson et al. 2011.

²⁶ Mobbs 2018.

²⁷ Mobbs, Hagan, et al. 2015.

²⁸ Mendl and Paul 2020.

²⁹ Mobbs, Wise, et al. 2021.

³⁰ Yoo, Hayden, and Pearson 2021.

Methods

Ethics information

This study has been reviewed and approved by the Institutional Review Board (IRB) of the California Institute of Technology. Subjects visiting the web app to participate in the study were provided with an informed consent document and were unable to proceed with the study until they had indicated consent via an anonymous electronic signature. All subjects were compensated at a base rate of \$10.00 per hour, with a possibility to earn a maximum of \$10.00 in bonus payments depending on their performance during the task.

Task Environment

The task environment takes the form of a browser-based application with a set of surveys followed by a sequence of trials, each trial being one randomized condition of several possible experimental conditions of a video game task.

On each trial, individuals in this environment must move their character (referred to as a *sprite*) around capturing rewards at foraging patches distributed around a central refuge, sometimes foraging alongside a computer-controlled AI representing a social other which they are told is controlled by a real person. The social other will vary in speed, being sometimes slower than the subject, sometimes faster. The speed of the social other is indicated by both the color and shape of the social other sprite on screen and these conventions are explicitly told to the subject at the beginning of the task. On some trials, a predator will appear and attempt to capture the participant or the social other, depending on whichever is closer at any given moment. The pairs of predator and social other speeds (their “fitness”) define each of the experimental conditions. There are eight possible conditions:

1. Slow Other, Fast Predator
2. Slow Other, Slow Predator
3. Slow Other, No Predator
4. Fast Other, Fast Predator
5. Fast Other, Slow Predator
6. Fast Other, No Predator
7. No Other, Slow Predator

<https://social-safety-task-1333a.web.app/>

8. No Other, Fast Predator

When the subject collects a token in a foraging patch, the coin will disappear for two seconds before respawning to prevent the player from staying in a single patch for the entire trial. Tokens collected by the social other do not disappear and can still be collected by the subject; this prevents the subject from seeing the social other as a competitor. If captured before the social other is captured or before reaching refuge on a single trial, the participant will lose the reward they have accumulated during that trial. Trials with no predator end after a uniformly distributed random time between 10 and 40 seconds, while trials with a predator can end in one of three ways:

1. The subject sprite is captured by the predator.
2. The subject sprite reaches the refuge.
3. The social other sprite is captured by the predator.

These trial-end states and the trial progression are illustrated in Figure 1. A screenshot of the virtual ecology labeled with the salient features of the task can be found in Figure 2. Trials with a predator present are broken into three phases according to the *threat imminence continuum* model.³¹ The model stipulates that organisms exhibit different adaptive behavioral repertoires and responses to threats depending on the likelihood of danger: in the *pre-encounter* phase, a threat has not yet appeared and the organism may be focused on preemptive measures; in the *post-encounter* phase, the organism has detected a potential threat; in the *circa-strike* phase, the organism is under attack by the threat and is engaged in active escape and avoidance behaviors.

³¹ Fanselow and Lester 1988.

The pre-encounter phase of a trial has a uniformly distributed random duration between two to 16 seconds during which no predator sprite appears on screen. The post-encounter phase also has a uniformly distributed random duration between two to 16 seconds during which the predator appears on screen at a random location on the edge of the arena and oscillates around the edge according to a random walk. In the circa-strike phase, the predator leaves the edge of the arena and attacks the players, inducing the social other sprite to move toward the refuge. After the trial ends, a message is displayed to the subject indicating the outcome of the trial. Before a predator trial begins, subjects are asked to perform a metacognition task, using a continuous slider with their mouse to indicate their confidence that they will escape on the next round given the type of social other and predator that will be present.

A participant in this task will have essentially two strategies available for successful performance of the task: remain near the refuge,

or exploit the decreased speed of the social other during “slow other” trials, effectively using them as a source of safety by sacrificing them to the predator. The task dynamics enable the investigation of which strategy is being made by subjects in the arena.

To make the claim that subjects were playing the game alongside a real person during the social other trials believable, the movement of the social other sprite in the arena needed to resemble that of the subject's own sprite. Because the social other sprite also needed to forage for tokens in the patches around the arena and escape from the predator when attacked, movement needed to be goal-directed. A natural choice of movement algorithm for the social other sprite would be the traditional “boid” algorithm,³² but this algorithm re-

³² Reynolds 1987.

Design

This study employs a within-subjects design in which each subject is exposed to all predator-other conditions. Counterbalancing of the experimental conditions was achieved by presenting conditions to subjects in a pseudo-random de Bruijn sequence.³³ This sequence results in every subject playing eight rounds of each experimental condition, for a total of 64 rounds. The study follows a 3×3 design where each experimental condition is some combination of predator and social other speeds as indicated in the diagram in Figure 3. The No Other / No Predator condition is not presented to the subjects since it is not a necessary control and so as to limit the amount of time subjects must spend performing the task.

³³ Aguirre, Mattar, and Magis-Weinberg 2011.

Sampling

Only data from subjects who are fluent in English, between 18 and 65 years of age, and have at least a 90% rating on Prolific was included. Prolific maintains a tool to set inclusion criteria for each study, ensuring that only those Prolific users who meet all criteria can participate. Prolific also maintains a messaging service to enable anonymous

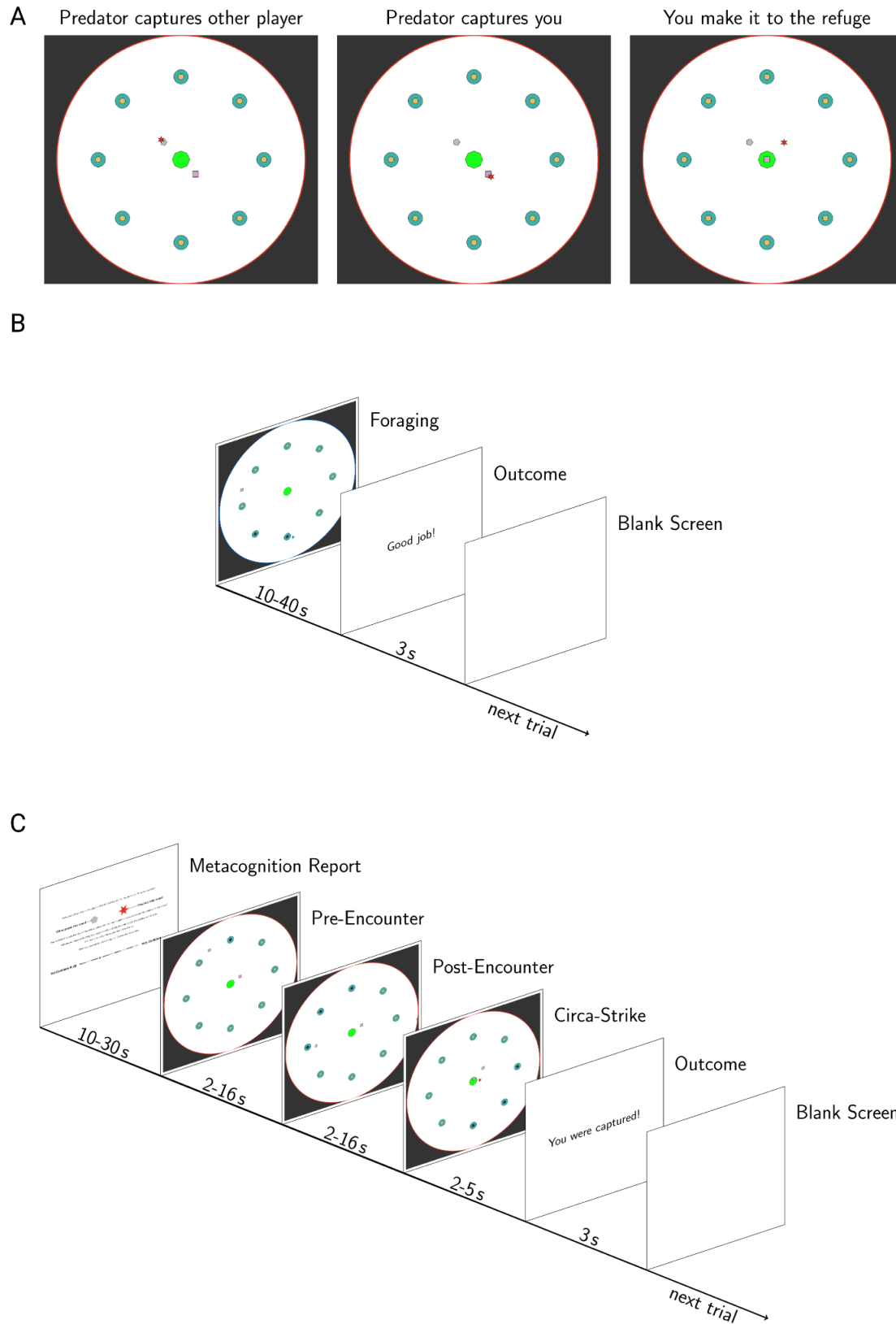


Figure 1: (A) Three ways for a trial in which a predator is present to end. (B) The trial progression for a trial in which no predator is present. (C) The trial progression for a trial in which a predator is present.

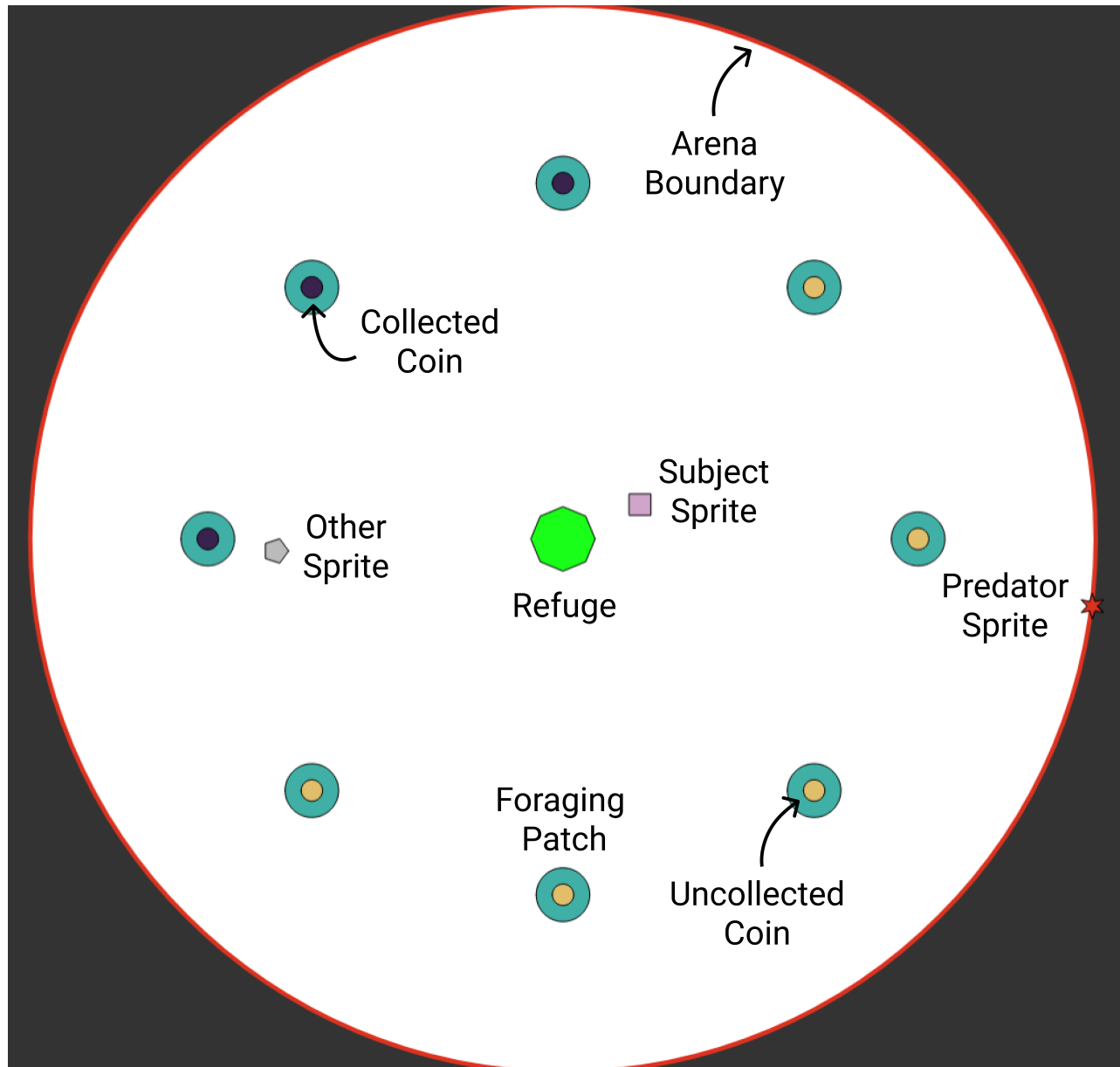
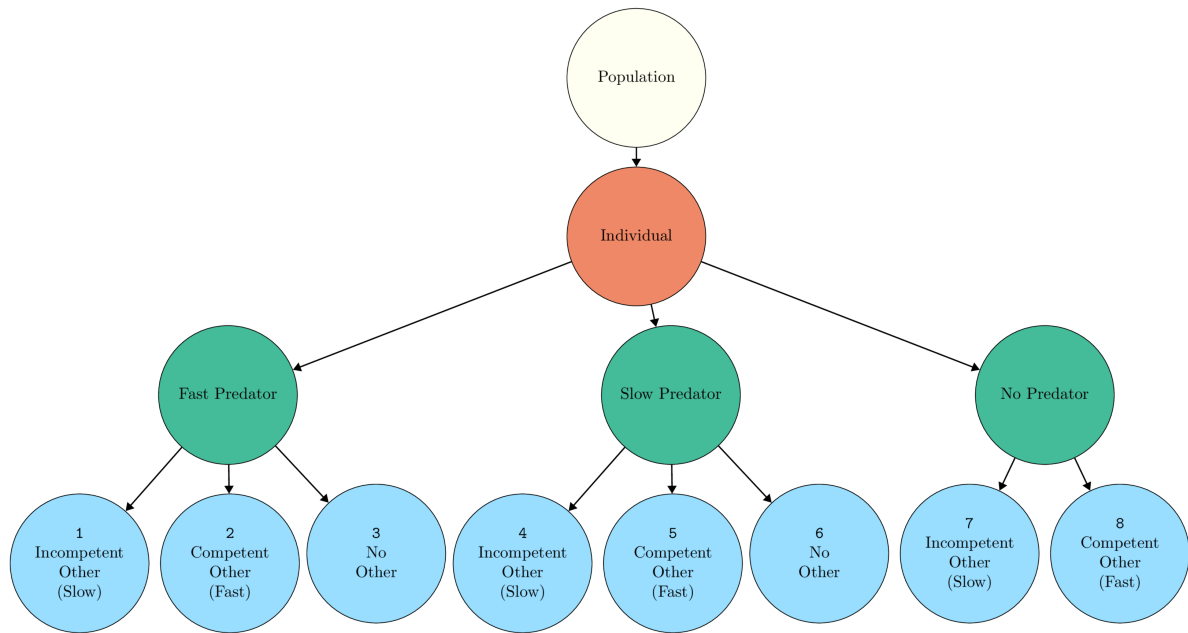
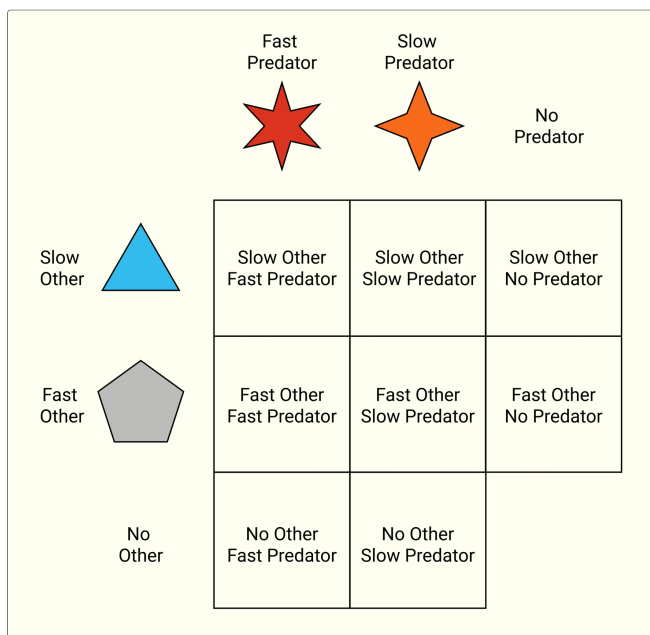


Figure 2: Still shot from the task illustrating the visual appearance of the arena. Labels indicate the salient features of the task environment.

A



B



C

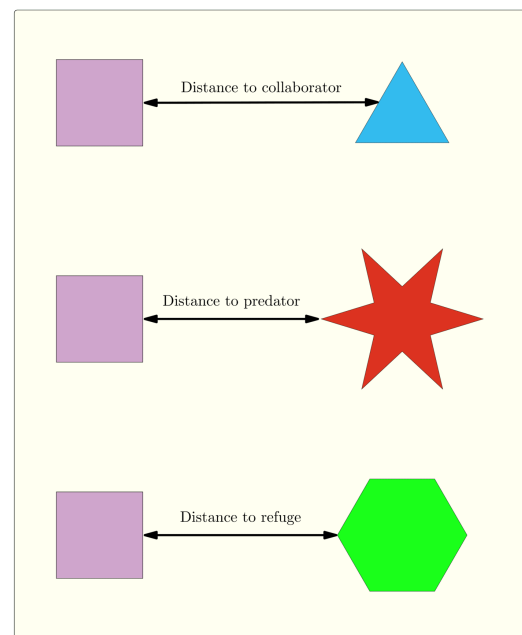


Figure 3: The levels of the experimental conditions. Subjects are pulled from a general population and exposed to eight total predator-other conditions. The colors of each level correspond to the colors of each level in the plate diagram in Figure 4.

Algorithm 1: Human-like AI control of social other sprite

Data: current sprite position vector $\mathbf{x}(t)$, current sprite velocity vector $\mathbf{v}(t)$, target position vector $\mathbf{g}(t)$, array of arrow-key direction vectors $\mathbf{D}$, transition probability $0 \leq p \leq 1$

Result: updated sprite position vector $\mathbf{x}(t+1)$, updated sprite velocity vector $\mathbf{v}(t+1)$

begin

$r \leftarrow \text{random}(0, 1)$ // uniform random number generation

if $r \leq p$ **then**

$\mathbf{x}(t+1) \leftarrow \mathbf{x}(t) + \mathbf{v}(t)$

$\mathbf{v}(t+1) \leftarrow \mathbf{v}(t)$

else

$\mathbf{h} \leftarrow \frac{\mathbf{g}(t) - \mathbf{x}(t)}{\|\mathbf{g}(t) - \mathbf{x}(t)\|}$

$\theta \leftarrow \text{array}(4)$ // create array with four elements

for $i = 0$ **to** 3 **do**

$\theta[i] \leftarrow \cos^{-1}(\mathbf{h} \cdot \mathbf{D}[i])$ // get angles

end

$j \leftarrow \min(\theta)$ // index of minimum angle

$\mathbf{v}(t+1) \leftarrow \mathbf{D}[j]$

$\mathbf{x}(t+1) \leftarrow \mathbf{x}(t) + \mathbf{D}[j]$

end

end

communication between experimenters and subjects. Data from any subjects using this messaging service to report that they have encountered a technical error with the task was excluded from the analysis. Subjects reporting technical errors were asked to "return" their study submission, an action within Prolific's interface which allows another user to take their place. Data was also excluded from the analysis for any subjects who ended the study early or failed the attention test following presentation of the task instructions. All analyses shown here were performed with $n = 67$ subjects.

Analysis

The notion of a spectrum of decision making strategies for safety-seeking can be formalized by mapping that spectrum to a behavioral metric of decision-making. Then, we can fit a generative model of

that behavioral metric to empirical data to determine what decision-making strategy an individual is using. Different parameter sets of the generative model will concentrate the distribution along different regions of the spectrum corresponding to different strategies.

As discussed previously, we theorize that if a person under threat has access to only two sources of safety, their safety-seeking behavior will vary continuously along a spectrum of strategies whose extremes are defined by the person maintaining close proximity to one or the other safety cue. The best metric for determining an individual's location on this spectrum is a behavioral measure we will refer to as the *distance-difference* or *DD*. The *DD* on frame t of a trial is the difference between the distance from subject sprite to refuge and the distance from subject sprite to the social other sprite. Since different participants had different monitor sizes and refresh rates, all distance measures are normalized to the on-screen size of the arena and frames (game state update cycles) are used as the unit of time.

A generative model that nicely captures continuous variation of the *DD* between safety-seeking strategies is the four-parameter Beta distribution with the mode-concentration parameterization. The Beta distribution captures the probability density of a continuous variable with upper and lower bounds, making it convenient for our modeling purposes since the geometry of the arena means that the distance-difference metric is bounded between negative and positive unity, $-1 \leq DD \leq 1$. However, the two-parameter Beta distribution is only defined over the interval $[0, 1]$, so we require the four-parameter Beta distribution to extend the support of the Beta distribution to arbitrary lower bound γ and upper bound δ . In this case, we set $\gamma = -1$ and $\delta = 1$. Beta distributions are typically parameterized by the shape parameters α and β , but these parameters do not have an intuitive effect on the shape of the distribution that would allow for straightforward formalization of hypotheses. Thus, I will use the mode-concentration parameterization which specifies the Beta distribution in terms of its mode ω and its concentration κ . The concentration κ may be thought of intuitively as an "inverse variance": the larger κ is, the narrower the density is. This parameterization only exists for so-called "concave" Beta distributions, which just means that the density goes to zero at the boundaries of its support. This is valid in the current context since we would not expect the distance-difference metric to regularly take values along its upper or lower limit which would require either one or both of the subject and other sprites to be at the edges of the arena, and unlikely scenario.

Here, I model the distance-difference observations in a fully Bayesian context. The model is hierarchical to account for population-level influences on uncertainty during decision-making. The plate

diagram of the model as well as an illustration of how the aforementioned spectrum of safety decisions maps onto the likelihood can be found in Figure 4. The basic premise is that if subjects are exploiting the decreased relative fitness of the social other for safety in a selfish-herd manner, then we would expect the mode of the Beta distribution of the distance-differences to be clustered on negative values for trials with a slow other, whereas the mode would be concentrated on positive values if subjects are staying close the refuge, which is what we would expect on trials with a fast social other. The formal specification of the hierarchical model is

Likelihood (Observations)

$$d_{nj} \sim \mathcal{B}(\alpha_j, \beta_j, \gamma, \delta) \quad (1)$$

$$\alpha_j = \omega_j(\kappa_j - 2) + 1 \quad (2)$$

$$\beta_j = (1 - \omega_j)(\kappa_j - 2) + 1 \quad (3)$$

$$\gamma = -1 \quad (4)$$

$$\delta = 1 \quad (5)$$

Priors (Experimental Condition)

$$\omega_j \sim \mathcal{C}(\mu_j^\omega, \sigma_j^\omega) \quad (6)$$

$$\mu^\omega = [-0.6, -0.3, 0, 0, 0.3, 0.6]^\top \quad (7)$$

$$\kappa_j \sim \mathcal{LN}(\mu_j^\kappa, \sigma_j^\kappa) \quad (8)$$

Hyperpriors (Population)

$$\sigma_j^\omega \sim \mathcal{LN}(\theta^{\sigma, \omega}, \phi^{\sigma, \omega}) \quad (9)$$

$$\mu_j^\kappa \sim \mathcal{N}(\theta^{\mu, \kappa}, \phi^{\mu, \kappa}) \quad (10)$$

$$\sigma_j^\kappa \sim \mathcal{LN}(\theta^{\sigma, \kappa}, \phi^{\sigma, \kappa}) \quad (11)$$

$$\theta^{\sigma, \omega} = 0 \quad (12)$$

$$\theta^{\mu, \kappa} = -1 \quad (13)$$

$$\theta^{\sigma, \kappa} = -1 \quad (14)$$

$$\phi^{\sigma, \omega} = 0.2 \quad (15)$$

$$\phi^{\mu, \kappa} = 0.2 \quad (16)$$

$$\phi^{\sigma, \kappa} = 0.2 \quad (17)$$

where $\mathcal{B}$ represents the Beta distribution, $\mathcal{C}$ represents the Cauchy distribution, $\mathcal{N}$ represents the Normal distribution, and $\mathcal{LN}$ represents the Log-Normal distribution. The integer n indexes the number of distance-difference observations within an experimental condition and the integer j indexes the experimental condition itself. For clarity, the indices map to conditions as follows:

- 1 $\rightarrow$ Slow Other, Fast Predator
- 2 $\rightarrow$ Slow Other, Slow Predator
- 3 $\rightarrow$ Slow Other, No Predator
- 4 $\rightarrow$ Fast Other, No Predator
- 5 $\rightarrow$ Fast Other, Slow Predator
- 6 $\rightarrow$ Fast Other, Fast Predator

The choice of priors can be justified as follows. The mode of the Beta likelihood, ω_j , receives a Cauchy prior as a formalization of the hypothesis that different experimental conditions will be characterized by decision-making strategies at different locations on the safety decision spectrum. Specifically, the chosen values of the μ_j^ω are chosen so that with high likelihood,

$$\omega_1 < \omega_2 < \omega_3 = 0 = \omega_4 < \omega_5 < \omega_6$$

Log-Normal priors are chosen for variables such as the κ_j which are positive-definite, with hyperparameters chosen to allow for diffuse distributions, reflecting the lack of knowledge about the uncertainty present in these decision-making strategies.

Although the ω, κ parameterization enables clear and interpretable quantification of my hypotheses, most statistical computing packages specify the Beta distribution using the traditional α, β parameterization, so I have chosen to transform the ω, κ parameters into the α, β parameters for ease of use in a computing context.

The model is specified in the probabilistic programming language Stan with outputs of the inference algorithms accessed via the CmdStanPy Python interface.³⁴ Initially, I attempted to fit the models using Hamiltonian Monte Carlo, an algorithm which is guaranteed to converge to the true posterior, but the computational resources required proved too great even for a High Performance Computing cluster. This high computational cost necessitated the use of approximation algorithms. Here, I employed meanfield variational Bayesian inference to fit the generative model.³⁵ The main idea behind variational inference is to approximate the posterior $g(\xi \mid \mathbf{D})$ over parameters ξ and observations $\mathbf{D}$ by another more computationally tractable distribution $q(\xi)$, with the goal of minimizing an upper bound on the Kullback-Leibler divergence between the approximate posterior and the true posterior. The primary assumption of the meanfield approximation is that the parameters are independent of each other so that the distribution $q(\xi)$ has the form

³⁴ Carpenter et al. 2017.

³⁵ Fox and Roberts 2012.

$$q(\xi) = \prod_{k=1}^M q(\xi_k) \quad (18)$$

The Stan language specifically employs the automatic differentiation variational inference (ADVI) algorithm for meanfield variational inference.³⁶

³⁶ Kucukelbir et al. 2015.

Results

Exemplar spatial movement trajectories in the arena as well as distance time series during trials of each condition can be seen in Figures 6 and 7. The initial results do not give credence to the hypotheses developed *a priori*. As can be seen in Figure 5, the empirical cumulative distribution functions (ECDFs) of the distance-difference metric are nearly identical when observations within each trial condition are pooled across all three phases of threat imminence, suggesting no behavioral differences across conditions. As might be expected, differences between the experimental conditions emerge when distance-difference observations are split up by threat imminence phase. The differences that do emerge, however, are in direct opposition to the *a priori* hypotheses I have developed here.

Up to this point, I have considered the distance-difference metric at every frame of the trial, but perhaps a more relevant determiner of survival is the distance-difference immediately at the start of the circa-strike phase. We can simultaneously develop an understanding of the patterns of safety decisions across subjects as well as which safety decisions are most associated with success on the task by examining the distributions of the distance-difference at circa-strike on successful and unsuccessful trials.

These distributions are plotted for all experimental conditions in Figure 8. The distributions were constructed as follows:

1. Extract the distance-difference on the first frame of the circa-strike phase across all trials.
2. Split the set of distance-differences by whether or not the subject escaped or was caught (successful vs. unsuccessful trials).
3. Compute the kernel density estimates (KDEs) of the distance differences for successful and unsuccessful trials across all conditions. The bandwidth for the the KDEs is chosen using least-squares cross validation.

Across all conditions, the highest density of distance differences at circa-strike on successful trials is clustered around $DD \approx 0.5$, suggesting that the strategy of staying close to the refuge is associated

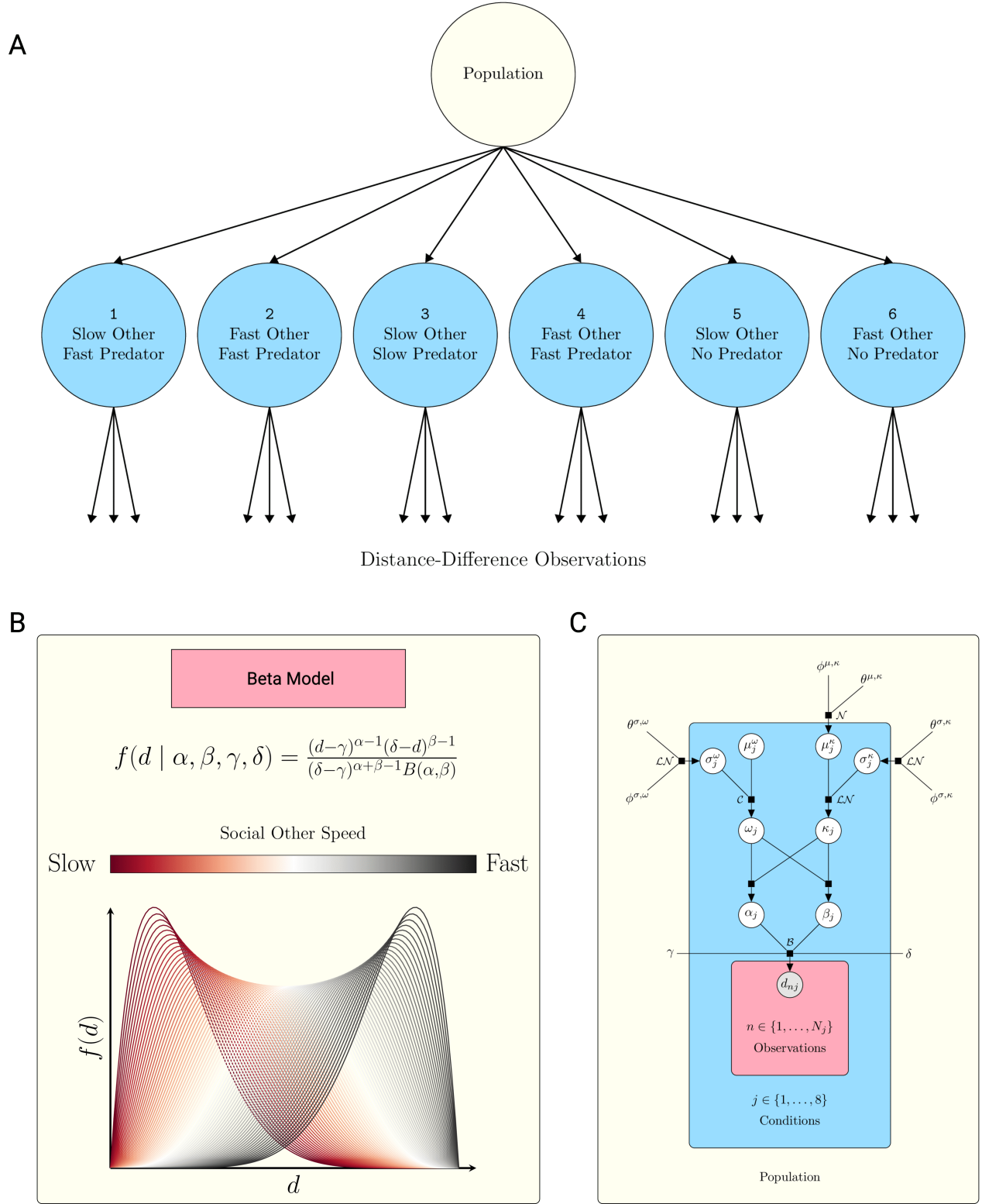


Figure 4: (A) The hierarchy of the model; population level influences impact the uncertainty (κ_j) of the distance-difference observations, but the modes ω_j are specific to each condition. (B) A schematic depiction of the spectrum of decision-making strategies under the four-parameter Beta model. (C) Plate diagram for the hierarchical model.

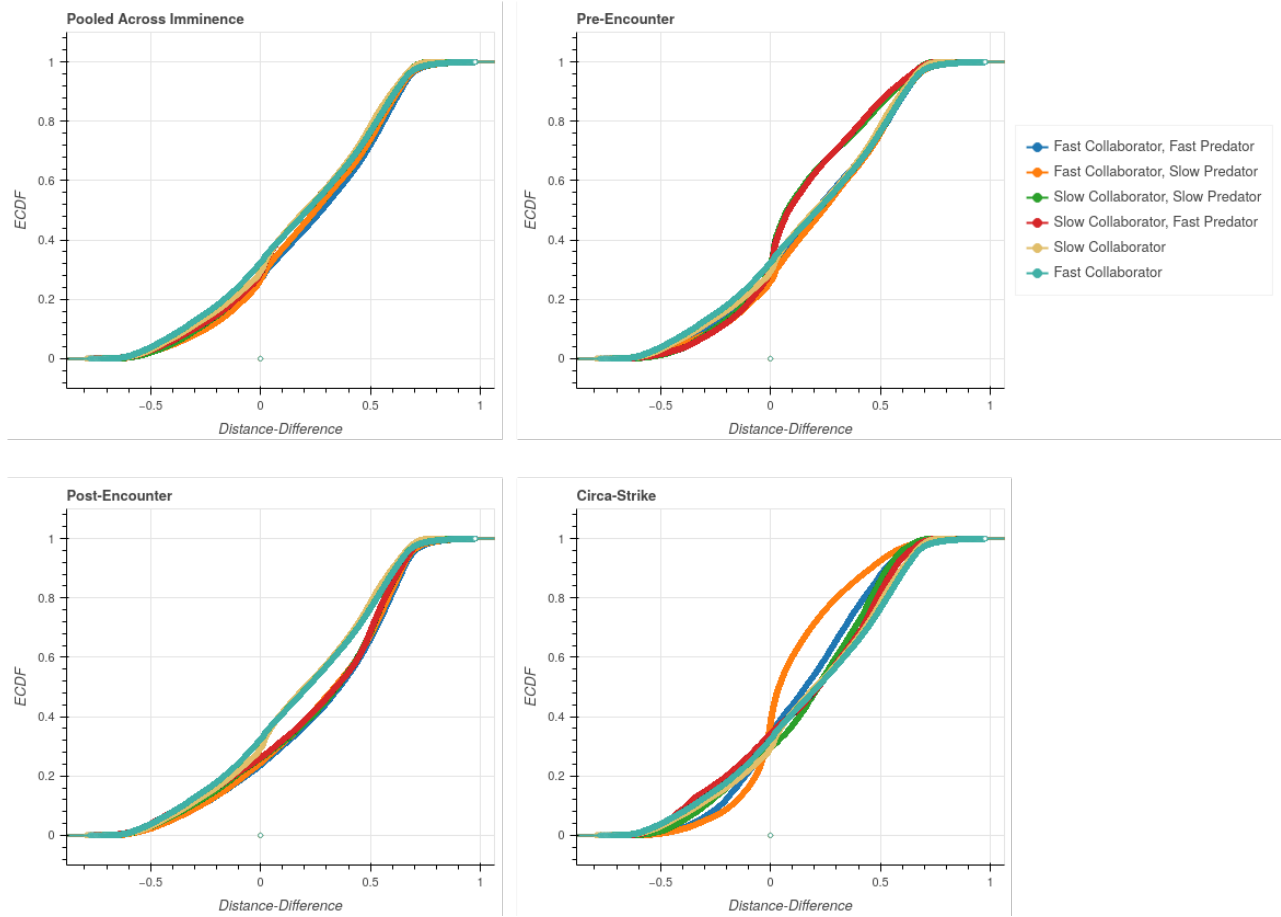


Figure 5: Empirical cumulative distribution functions of the distance-difference metric pooled across all threat imminence phases as well as within each of the three threat imminence phases. Only the six conditions in which a social other is present are shown since the distance-difference metric only exists in these conditions.

with the greatest likelihood of escape. This association is reinforced by the multimodal densities of the unsuccessful trials: though at least one mode of the KDEs for unsuccessful trials is in the positive interval across all conditions, there is another mode in the negative interval which is greater than the KDE for successful trials across the entire interval. This result indicates that the strategy of staying closer to the social other than to the refuge is more likely to lead to the subject getting caught by the predator. So not only does the selfish-herd style strategy appear to not be as viable as staying close to refuge, subjects are also more likely to stay closer to refuge than the social other regardless of the fitness of the social other or the predator.

The fit of the Bayesian generative model for the distance-differences obtained via variational inference can be assessed graphically using the empirical cumulative distribution functions of the distance-difference observations and the simulated distance-differences generated by the posterior predictive distribution of the model. Briefly, the

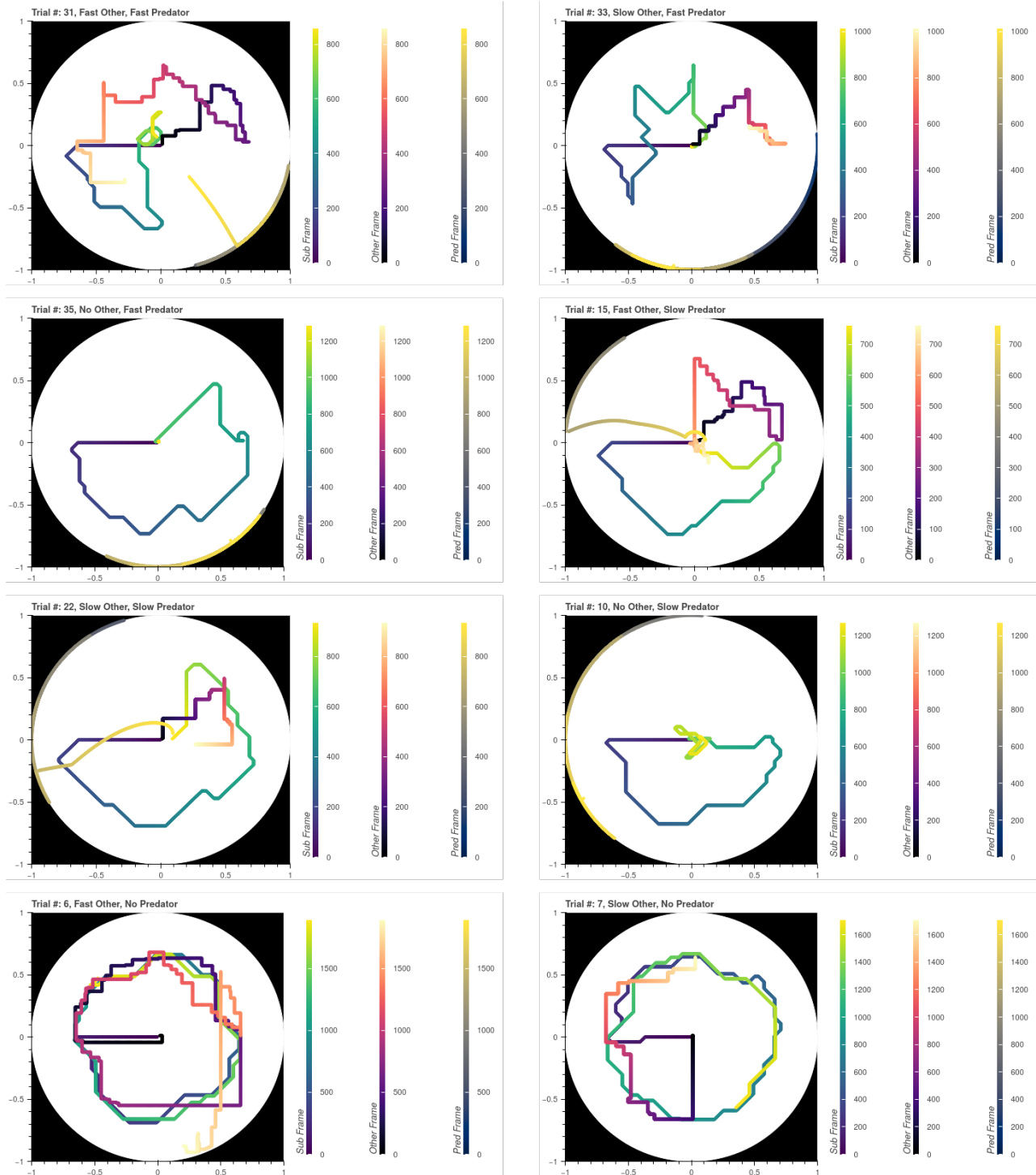


Figure 6: Depiction of the trajectories taken by the subject, social other, and predator around the arena during exemplar trials of each condition for a single subject. The colors indicate the frame number of the trajectory point.

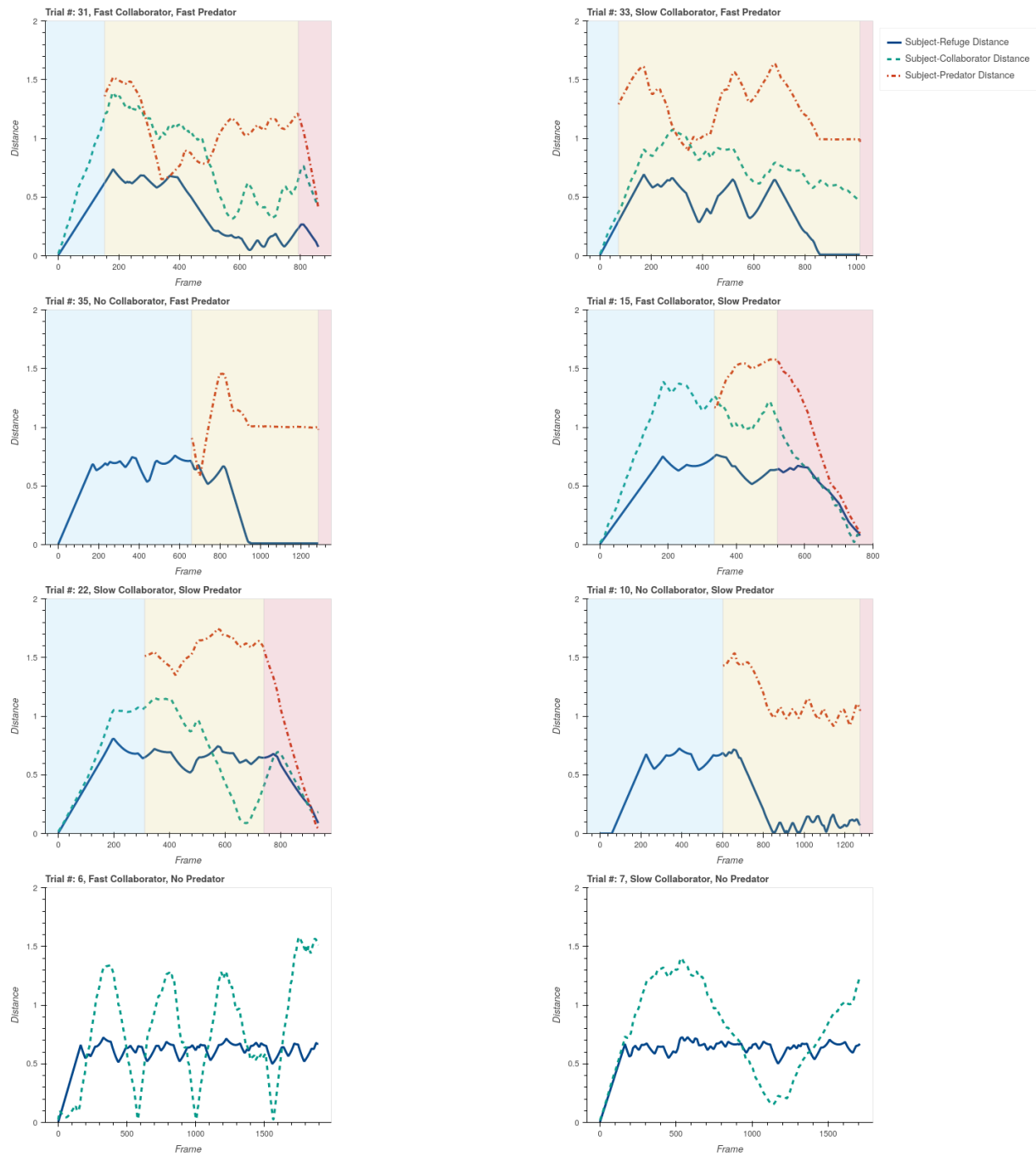


Figure 7: Time series of distances between subject and social other, predator, and refuge during exemplar trials of each condition for a single subject. The blue block indicates the frames corresponding to the pre-encounter phase, the yellow block indicates the post-encounter frames, and the red block indicates the circa-strike frames.

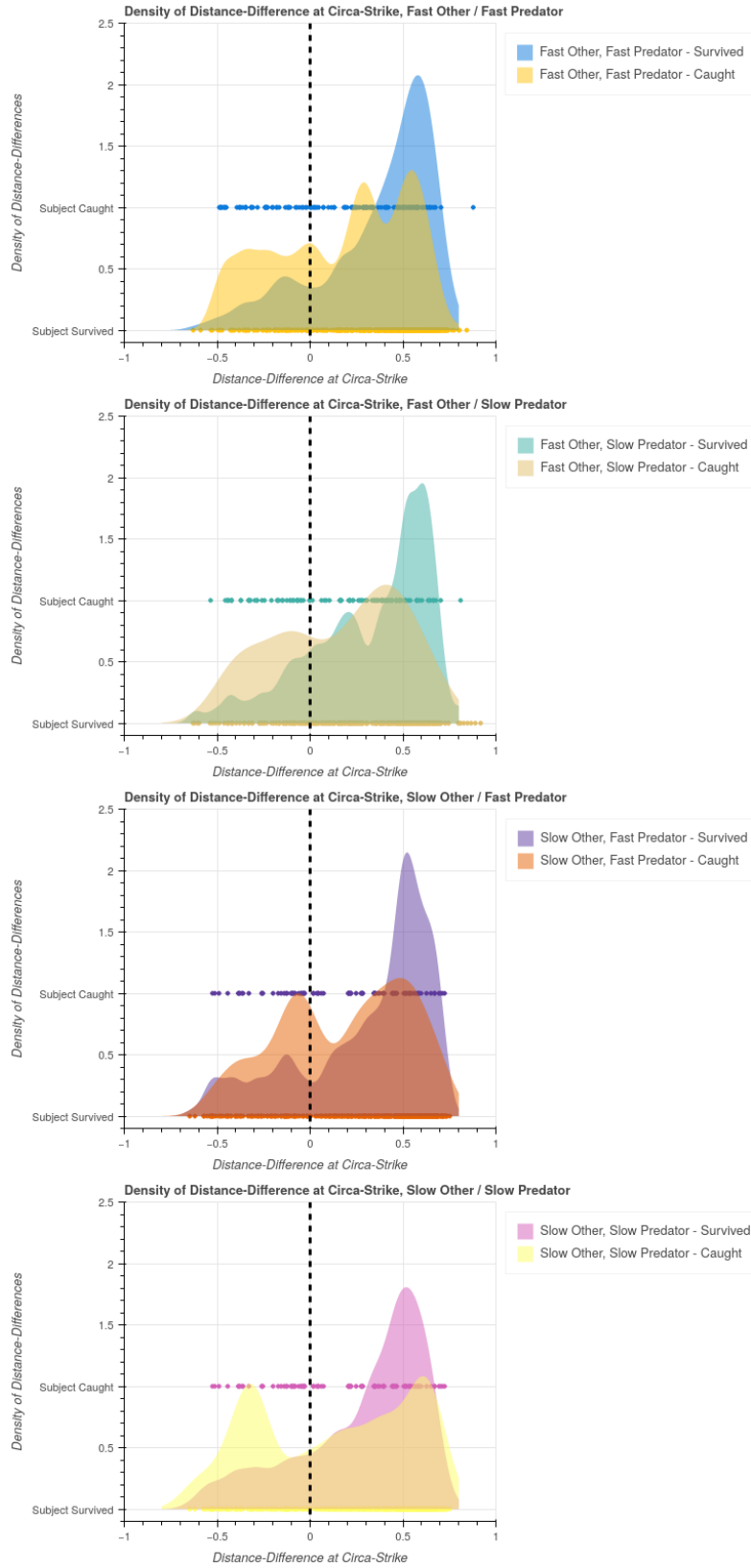


Figure 8: Kernel density estimates of the distance-difference distributions on both successful and unsuccessful trials across threat imminence phases.

posterior predictive distribution is obtained in two steps:

1. First, sample parameters from the approximate posterior distribution.
2. Plug these parameters into the the likelihood of the model and sample from this likelihood.

More formally, given a set of N observations of distance-differences $\mathbf{D} = \{d_1, \dots, d_N\}$ and a model with likelihood $f(\tilde{d} \mid \xi)$ and posterior $g(\xi \mid \mathbf{D})$, one defines the posterior predictive distribution as

$$p(\tilde{d} \mid \mathbf{D}) = \int_{\Xi} d\xi f(\tilde{d} \mid \xi) g(\xi \mid \mathbf{D}) \quad (19)$$

where $\tilde{d}$ represents a distance-difference value not taken from original set of observations $\mathbf{D}$ and Ξ is the entire parameter space of the parameter ξ . Essentially, comparing the empirical distribution of the observed distance-differences to the posterior predictive distribution of distance-differences addresses the question “How capable is my model of generating data sets similar to the data I have observed so far?”. This process is called is a “posterior predictive check” in the Bayesian inference literature.³⁷ To obtain a data set of simulated distance-differences and run a posterior predictive check, I perform the following steps:

³⁷ Gelman 2014.

1. Generate 4000 sets of 200 samples from the posterior predictive distribution for each experimental condition.
2. Use each data set to construct an ECDF.
3. Compute the quartiles of the ECDFs.
4. Overlay a plot of the ECDF of actually observed distance-differences on a plot of the posterior predictive ECDF quartiles.

This procedure allows one to visually check whether the measurements made during the task could have been generated by the model in question. If significant portions of the ECDF of the observed distance-differences lies outside the quartiles of the posterior predictive ECDFs, then the model may not be capable of generating the data. In this instance, we find that for all experimental conditions, small portions of the distance-difference ECDF lie outside the posterior predictive ECDF quartiles, implying that the variational fit is not perfect. Thus, we must take any inferences we make from this model with a grain of salt. Variational approximations are not expected to be perfect, and the general shape of the posterior predictive ECDFs is sufficiently well-matched to that of the actually observed ECDFs to give credence to the model.

As can be seen from the posterior predictive checks in Figure 9 and the marginal posterior plots in Figures 10 and 11, the marginal posteriors for the Beta modes ω_j across all conditions are centered on small positive values. This modeling result further supports the conclusion that, contrary to my predictions, people are not disposed to exploit weaker social others as a source of safety; at the very least, this particular experimental paradigm does not induce a preference for such a survival strategy.

Interestingly, the marginal posteriors for the concentration parameters κ_j indicate that there is little uncertainty in the decision-making strategies adopted by participants: they maintained distance-differences within a narrow range across all conditions. Of note, the conditions in which a slow predator was present (κ_2, κ_5) appear to be associated with slightly decreased decision-making uncertainty than the other conditions. This is likely because the slow speed of the predator made escape more predictable for subjects.

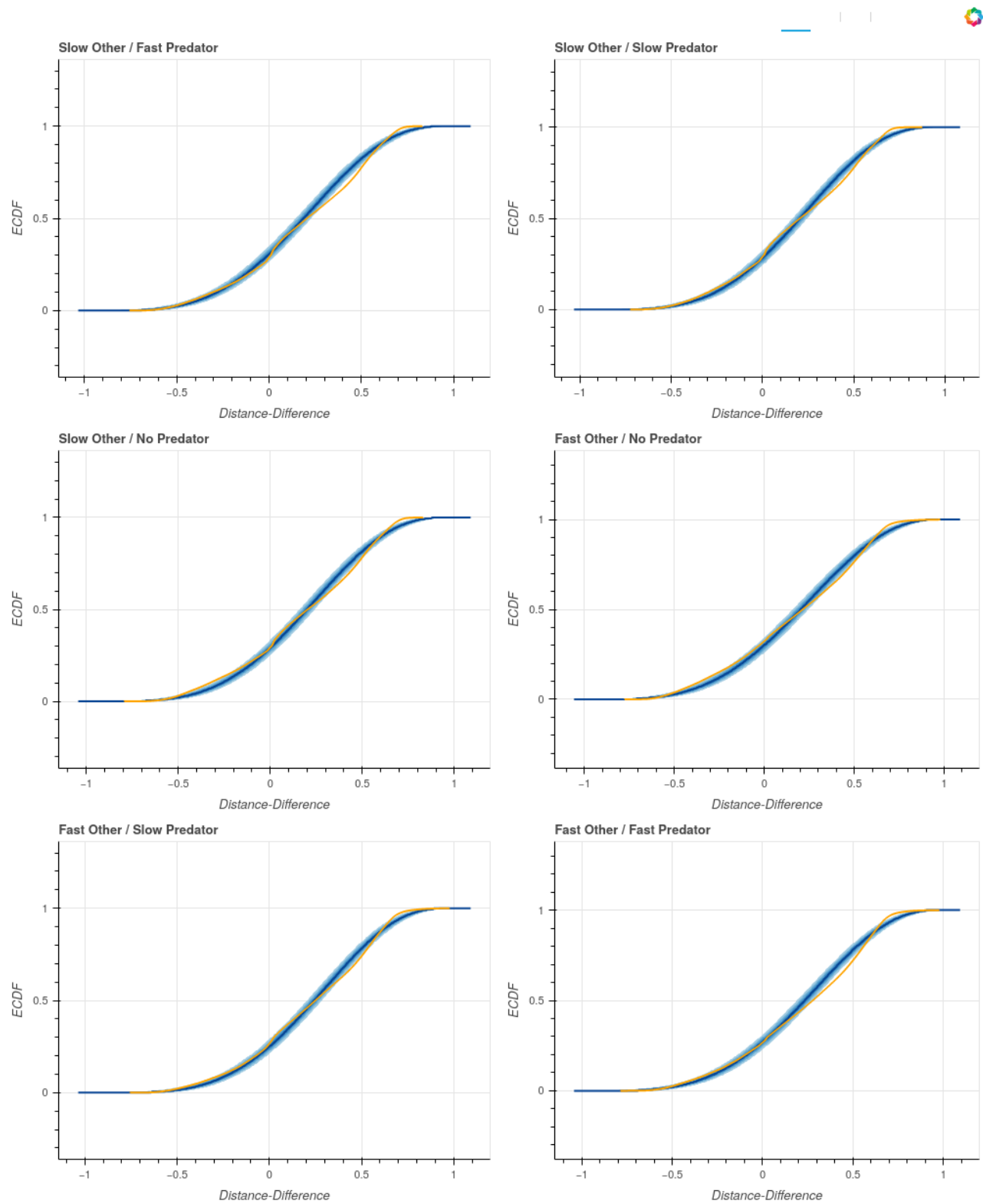


Figure 9: Posterior predictive checks showing the ECDF of observed distance-differences overlayed on top of the quartiles of the posterior predictive ECDFs.

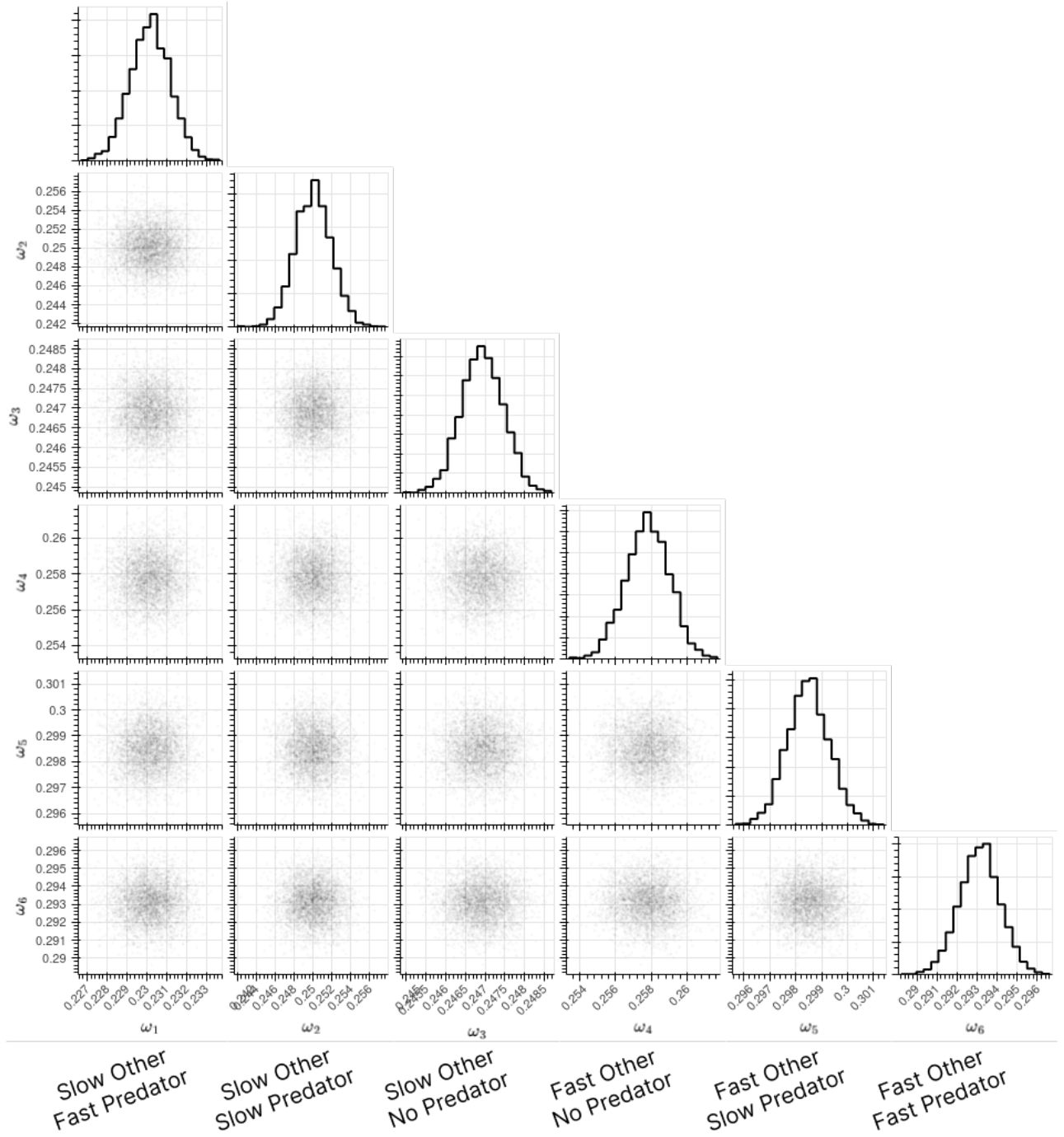


Figure 10: Corner plot illustrating the marginal posterior histograms and joint scatter plots of samples for the six mode parameters of the Beta model.

At the beginning of the task, subjects were asked to complete four surveys for measuring traits related to psychopathology and personality. These surveys included

1. *BIS/BAS Scales*:³⁸ A questionnaire for assessing an individual's dispositions with regards to the hypothesized behavioral inhibition system (BIS) governing avoidance and the behavioral activation system (BAS) governing approach, with subscales measuring drive, fun-seeking, and reward responsiveness within the BAS. Scored by summing responses.
2. *Depression Anxiety Stress Scales*:³⁹ A 42 item questionnaire with subscales measuring trait depression, trait anxiety, and trait stress in individuals. Scored by summing responses.
3. *Interpersonal Reactivity Index*:⁴⁰ A questionnaire for assessing an individual's capacity for empathy, with subscales for perspective-taking, empathic concern, tendency to experience distress when others are distressed, and imaginative skill. Scored by summing responses.
4. *Moral Foundations Sacredness Scale*:⁴¹ A questionnaire for assessing an individual's moral convictions by asking how much money it would take for them to violate those convictions. Can be scored by either taking the mean response or by counting the number of times a subject said they would never violate a conviction for any amount of money.

³⁸ Carver and White 1994.

³⁹ P. Lovibond and S. Lovibond 1995.

⁴⁰ Davis 1983.

⁴¹ Graham and Haidt 2012.

Of primary interest here are associations between the survey scores and behavioral metrics related to decision-making strategies. Three behavioral metrics were compared to the survey scores:

1. The total reward earned by the subject over the entire duration of the task. This is merely the sum of all tokens collected during the task.
2. The median distance-difference within a threat imminence phase across all trials and conditions.
3. The median reward rate within a threat imminence phase across all trials and conditions. If C_{TP} is the total number of tokens collected during threat imminence phase P of trial T with total number of frames F_{TP} , then the reward rate for that trial and threat imminence phase is computed as

$$\frac{C_{TP}}{F_{TP}} \quad (20)$$

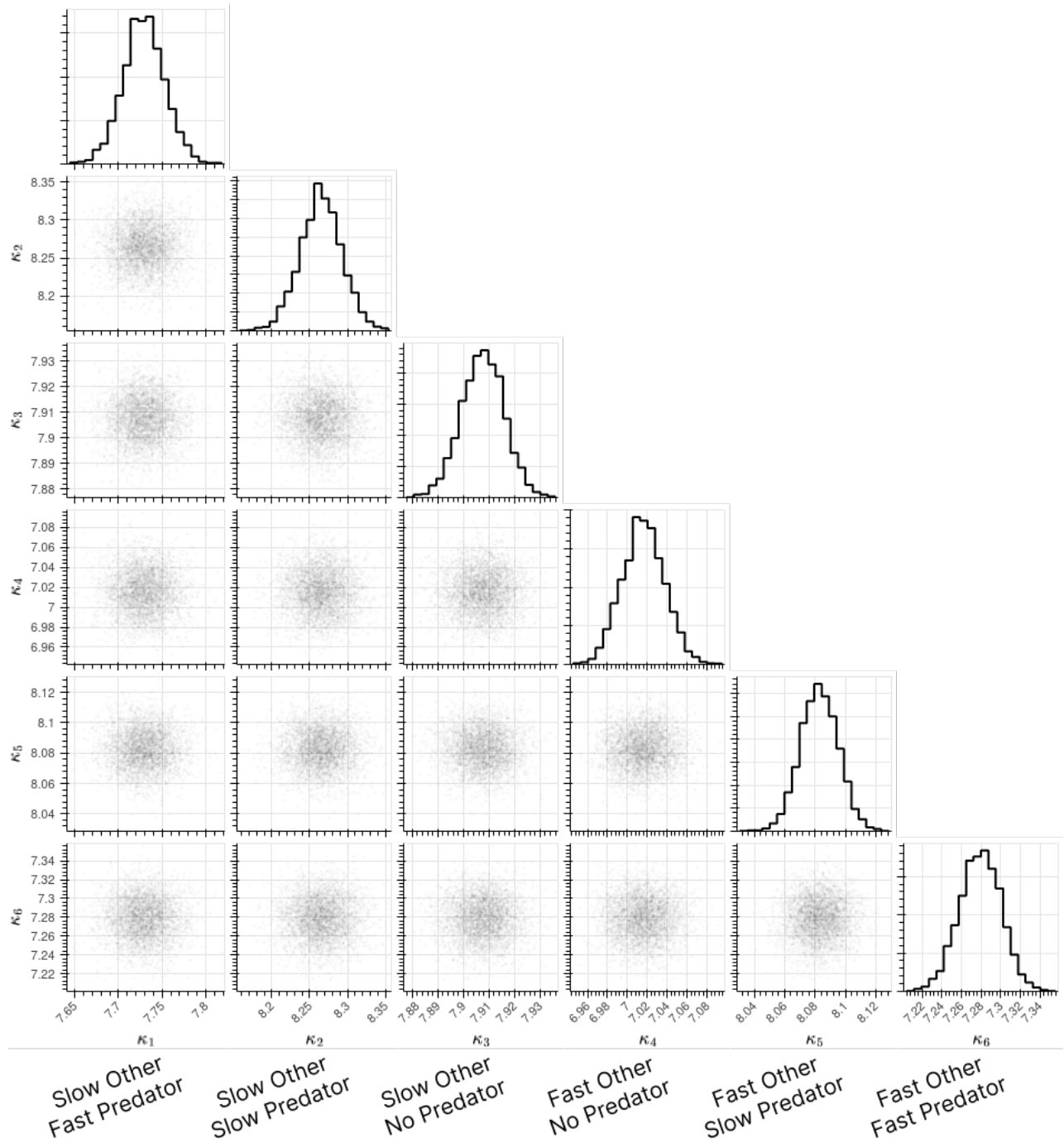


Figure 11: Corner plot illustrating the marginal posterior histograms and joint scatter plots of samples for the six concentration parameters of the Beta model.

Because the reward rate is often a very small number approaching zero (and sometimes equal to zero), I use the square root of the reward rate to transform the reward rate into more familiar ranges of numbers. This has no effect on the statistics since the square root is a monotonic transformation.

Though linear dependencies between two random variables may be assessed using metrics such as their Pearson correlation, I specifically wanted to account for the possibility of nonlinear dependencies between the survey scores and behavioral metrics. Therefore, pairwise associations were measured using approximations of the Maximal Information Coefficient (MIC). The idea behind the MIC is simple: the mutual information between two random variables serves as a general measure of the strength of the functional relationship between them, regardless of whether that relationship is linear or nonlinear.⁴² But approximating the mutual information of a pair of variables depends on how those variables are binned. Algorithms for computing the MIC optimize the binning scheme to determine the maximum possible value of the mutual information. Because the MIC is computationally expensive, two approximations have been developed, each with different strengths and weaknesses: the MIC_e , which trades statistical power for equitability (sensitivity to noise), and the TIC_e (Total Information Coefficient), which trades equitability for statistical power.⁴³ As recommended in the literature, I take a two-pronged approach, quantifying the statistical significance of pairwise associations between survey and behavioral metrics using the TIC_e , but using the MIC_e to report the strength of the associations. In all of the figures below, the MIC_e and the associated p -value derived from the TIC_e are reported on every pairwise plot. Computations of these metrics were performed with the MICtools Python toolbox.⁴⁴ Since many pairwise comparisons were performed, p -values were adjusted for multiple comparison using the Storey q -value to control the family wise error rate.⁴⁵

Only two associations emerged as statistically significant at the $p = 0.05$ level according to their TIC_e values:

1. The association between median distance difference during the circa-strike phase and BAS Drive score, with a slightly positive relationship between BAS Drive and the distance-difference. Though $p = 0.048$ (right on the border of significance), the $MIC_e = 0.215$ suggests that the association is weak.
2. The association between median reward rate during the pre-encounter phase and BAS Reward Responsiveness score. Though $p = 0.025$, the $MIC_e = 0.243$ suggests again that the association is weak.

⁴² D. N. Reshef et al. 2011.

⁴³ Y. A. Reshef et al. 2015.

⁴⁴ Albanese et al. 2018.

⁴⁵ Y. A. Reshef et al. 2015.

Discussion

One obvious critique of the analyses presented here is that data was collected from too few subjects, and that with a larger sample size, different patterns of safety decisions may emerge. However, we as neuroscientists and psychologists are interested not in small behavioral effects for their own sake; rather, we want to understand mechanisms of behavior that are clearly manifested across large populations. If a behavioral disposition is sufficiently widespread within a population, then as long as the experiments meant to elucidate them are well-designed and theory driven, it should not take a large sample size to bring those patterns to light.

It is conceivable that this virtual ecology as it has been designed is not appropriate for assessing the question “Do humans exploit each others’ weaknesses as a source of safety?”. Many possible variations on this task environment exist: we could increase the speed disparity between the subject sprite and the other sprites, randomly block the subject’s access to the safety refuge at different points during a trial, or (in accordance with the assumptions of selfish-herd theory) make the predator’s attack more instantaneous and stochastic. It may even be that no modifications could induce the hypothesized behaviors, and that to do so, the entire virtual ecology would need to be redesigned from the ground up. But I believe that the task environment is well suited to address this question for the simple reason that it possesses a critical element that would be required by any task attempting to investigate exploitative safety decisions: the presence of two potential safety cues (the social other and refuge) which enable us to distinguish which decision-making strategy subjects are *choosing* rather than *being forced* to take.

Possible improvements to the analyses here include the addition of further hierarchical models which explicitly model individual variability as well as distance-differences at the level of threat imminence phases within trial conditions. Though modeling individual variability may certainly reveal interesting patterns of decision-making in social ecologies, it is more doubtful that modeling at the level of threat imminence phases would align with the *a priori* hypotheses presented here. As indicated by clear patterns in their distributions, the distance-differences in all threat imminence phases are concentrated around positive values, giving credence to the notion that people tend to be risk averse and choose to remain near guaranteed sources of safety. This can be most clearly seen by visually comparing the prior distributions of the ω_j to their posterior distributions as is done in Figure 20.

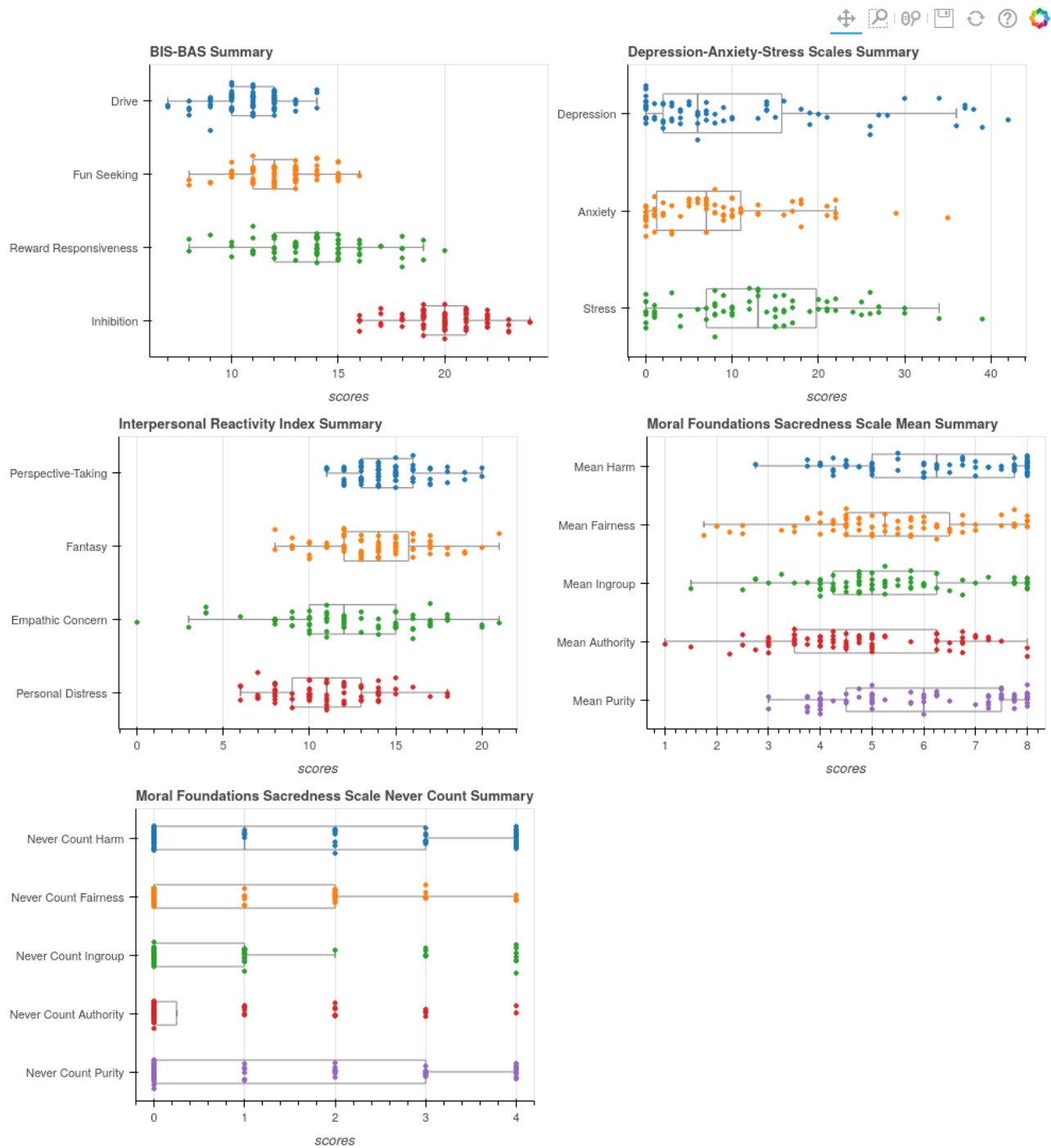


Figure 12: Visual strip plots summarizing the scores across all surveys and subscales.

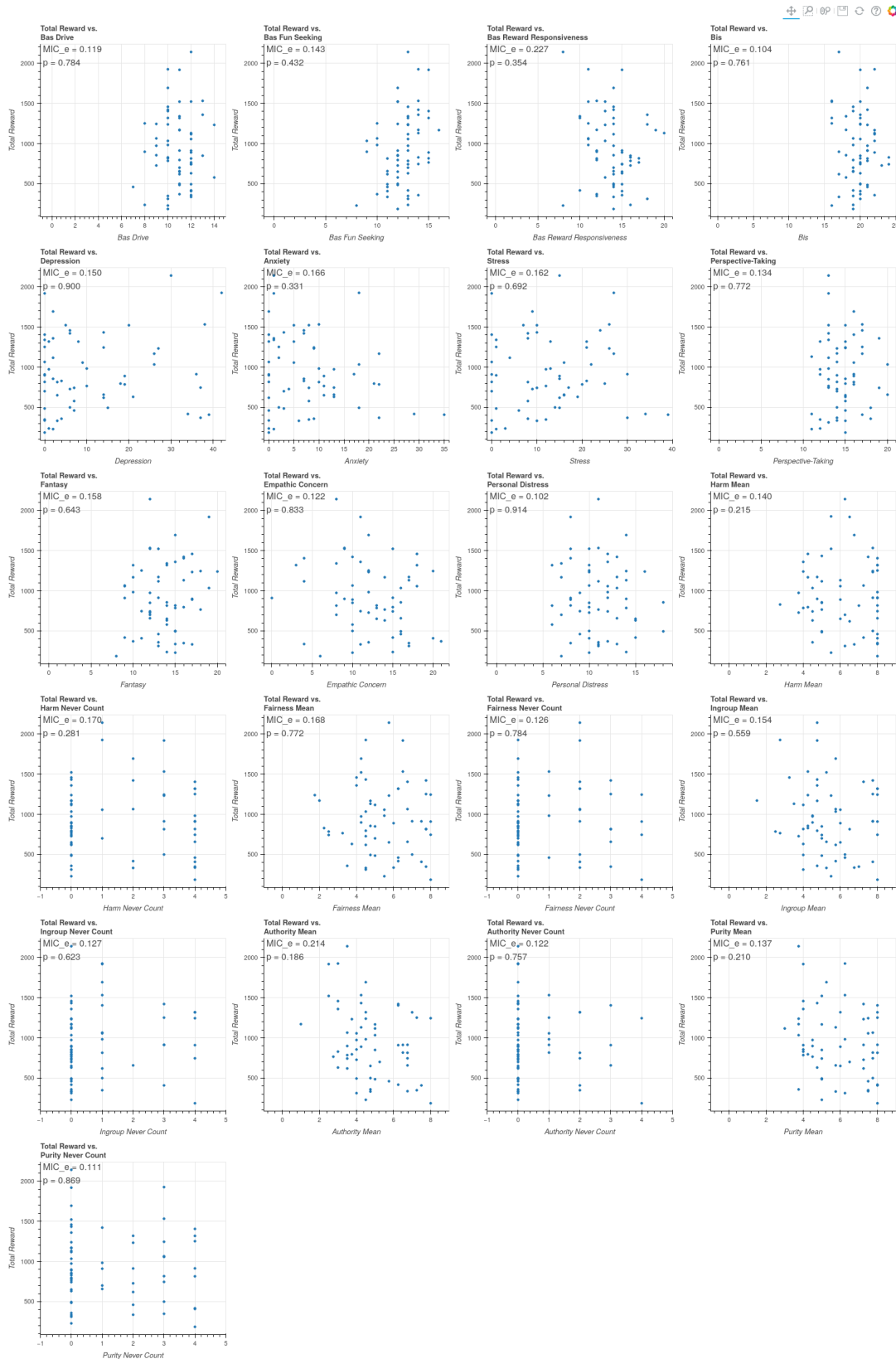


Figure 13: Total reward plotted against the survey scores. The MIC_e and the p-value are indicated in the top left corner.

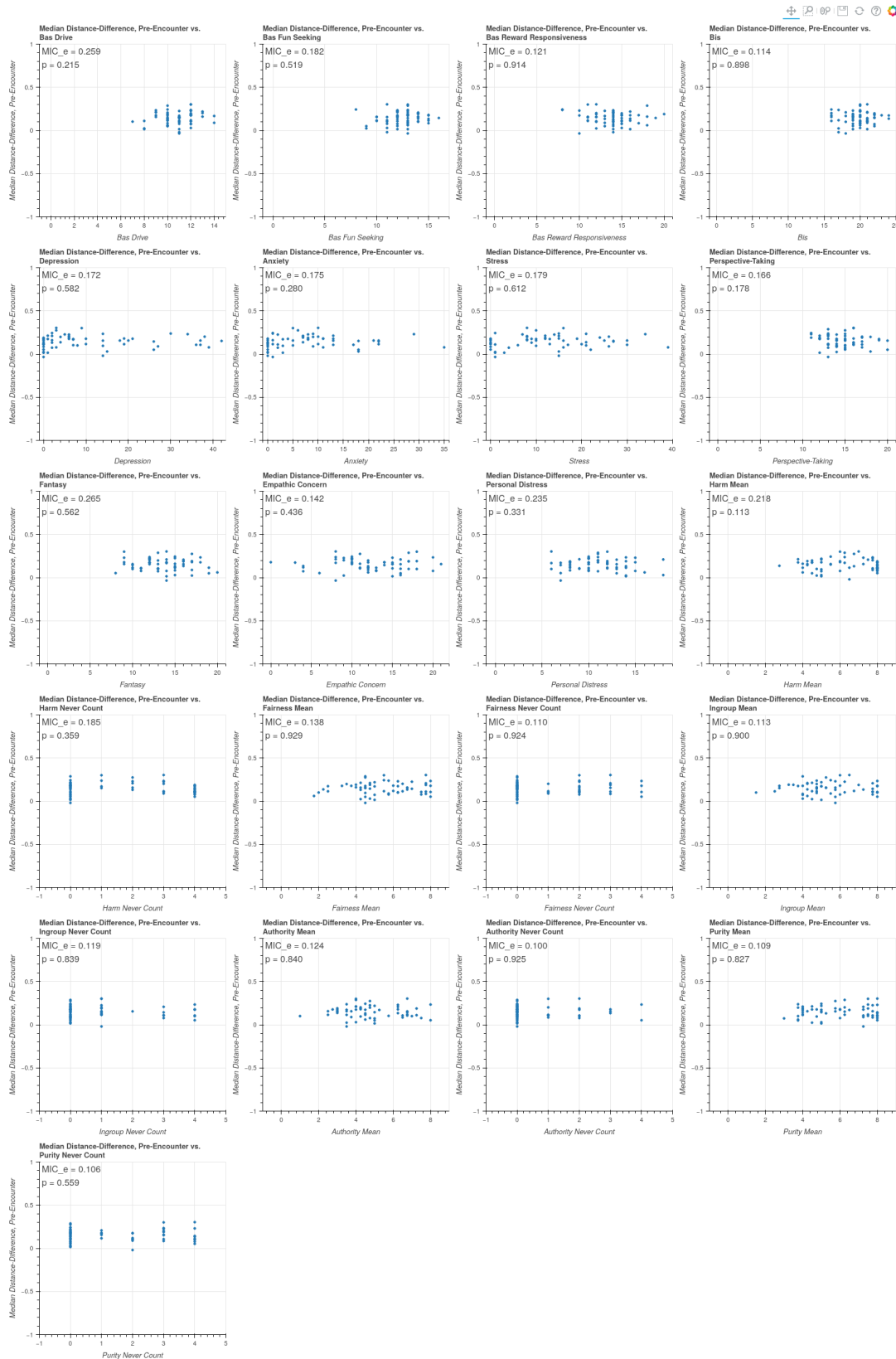


Figure 14: Median distance-difference during the pre-encounter phase plotted against the survey scores. The MIC_e and the p-value are indicated in the top left corner.

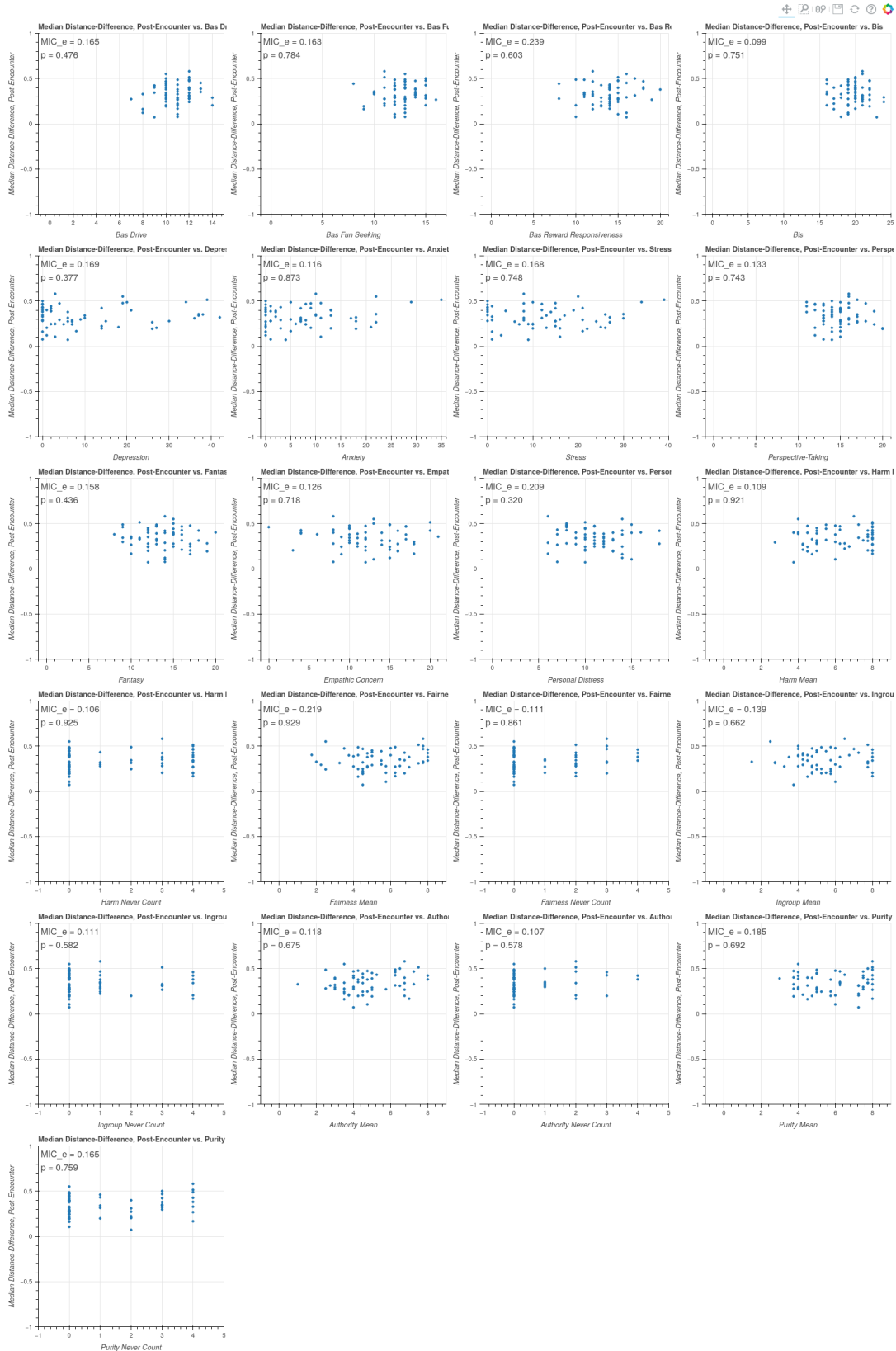


Figure 15: Median distance-difference during the post-encounter phase plotted against the survey scores. The MIC_e and the p-value are indicated in the top left corner.

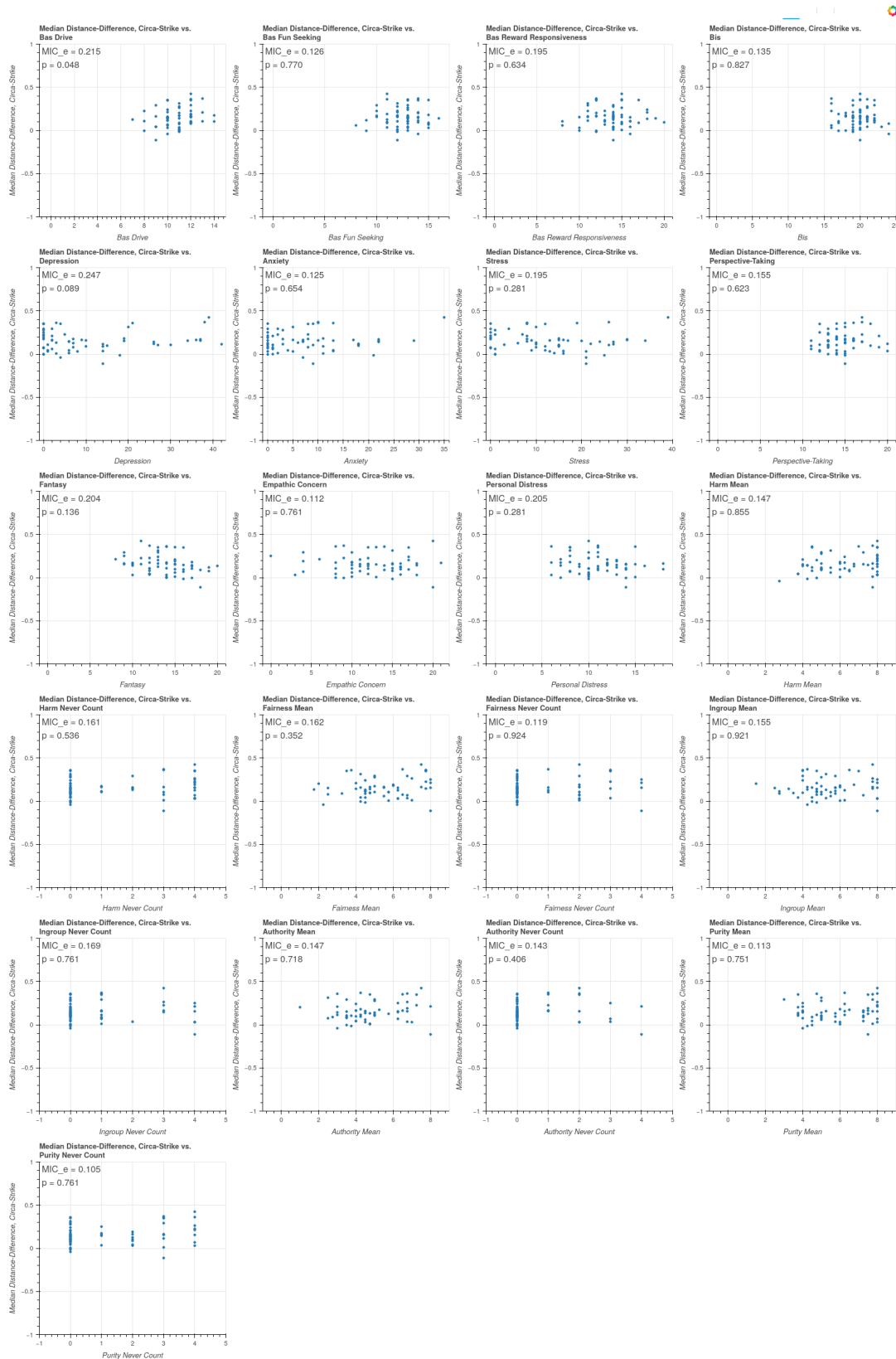


Figure 16: Median distance-difference during the circa-strike phase plotted against the survey scores. The MIC_e and the p-value are indicated in the top left corner.

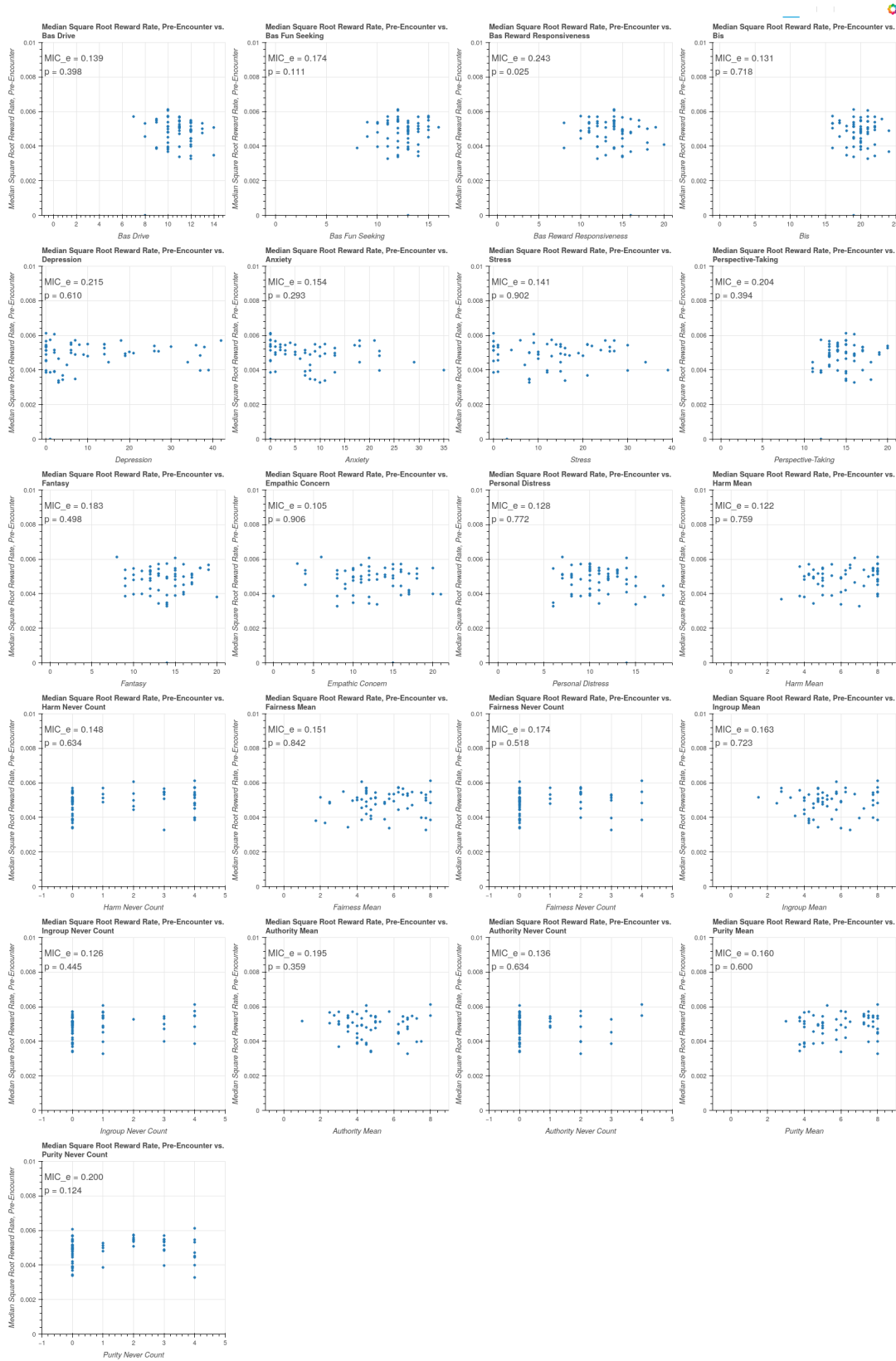


Figure 17: Median square root reward rate during the pre-encounter phase plotted against the survey scores. The MIC_e and the p-value are indicated in the top left corner.

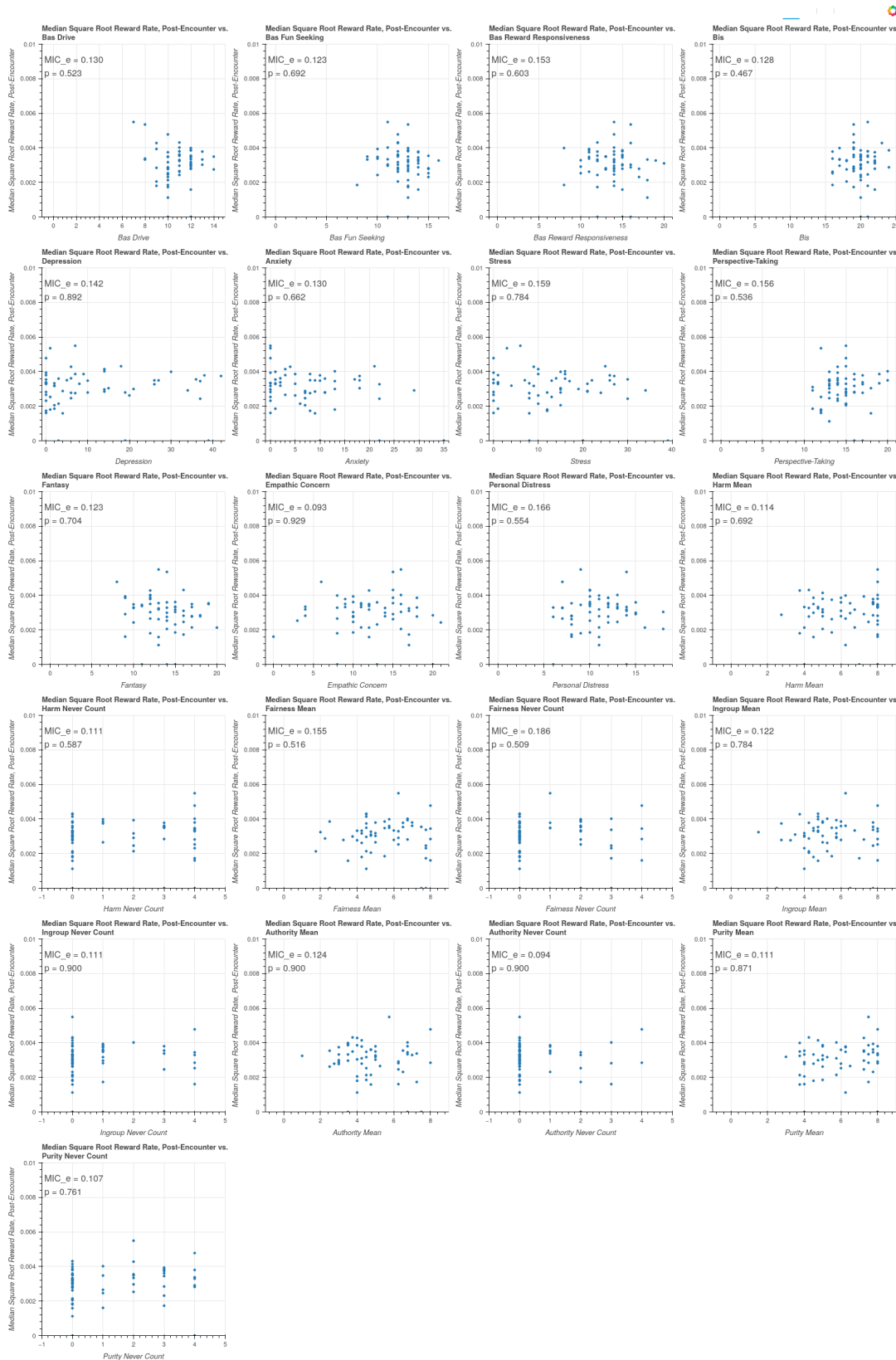


Figure 18: Median square root reward rate during the post-encounter phase plotted against the survey scores. The MIC_e and the p-value are indicated in the top left corner.

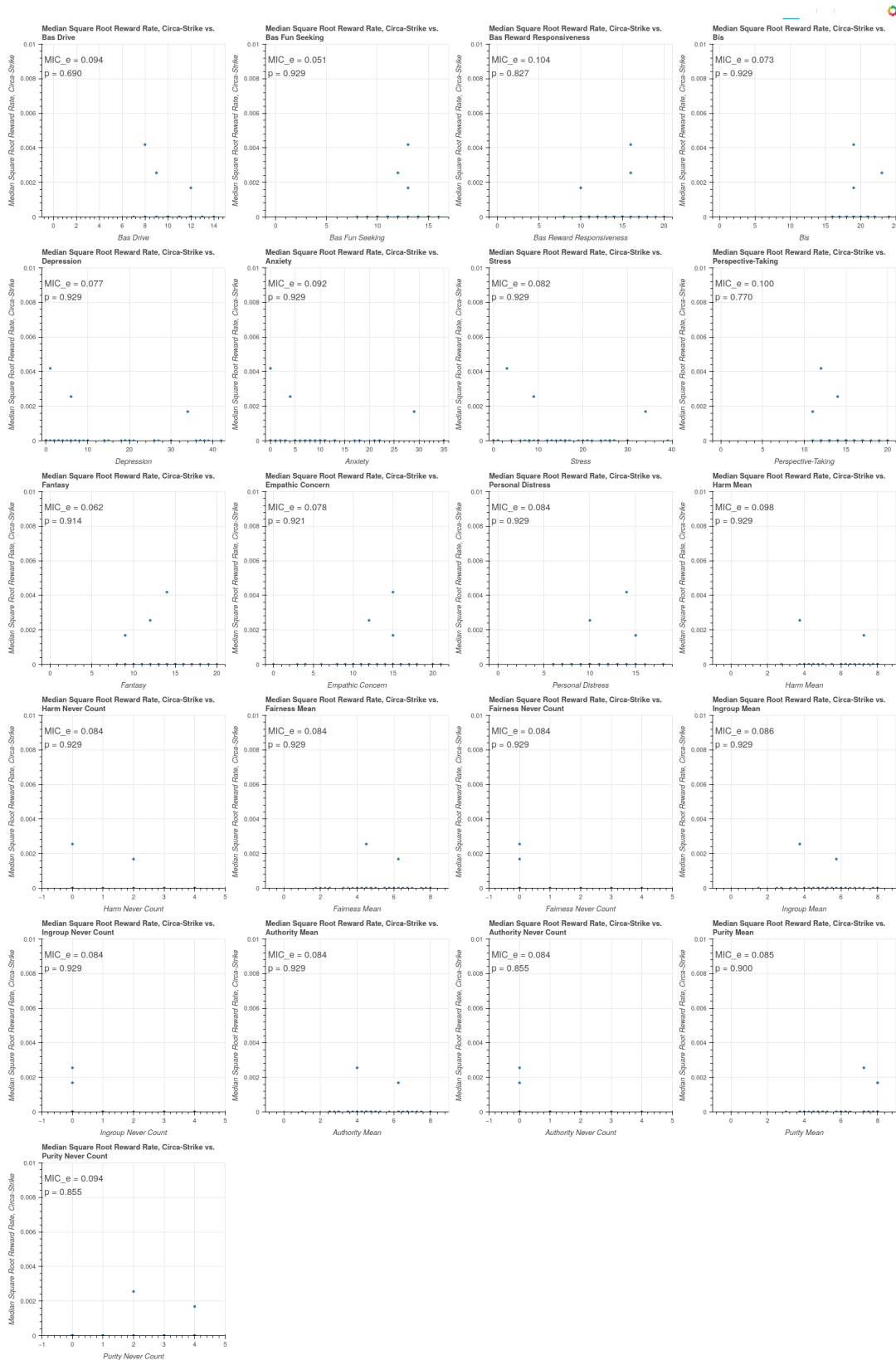


Figure 19: Median square root reward rate during the circa-strike phase plotted against the survey scores. The MIC_e and the p-value are indicated in the top left corner.

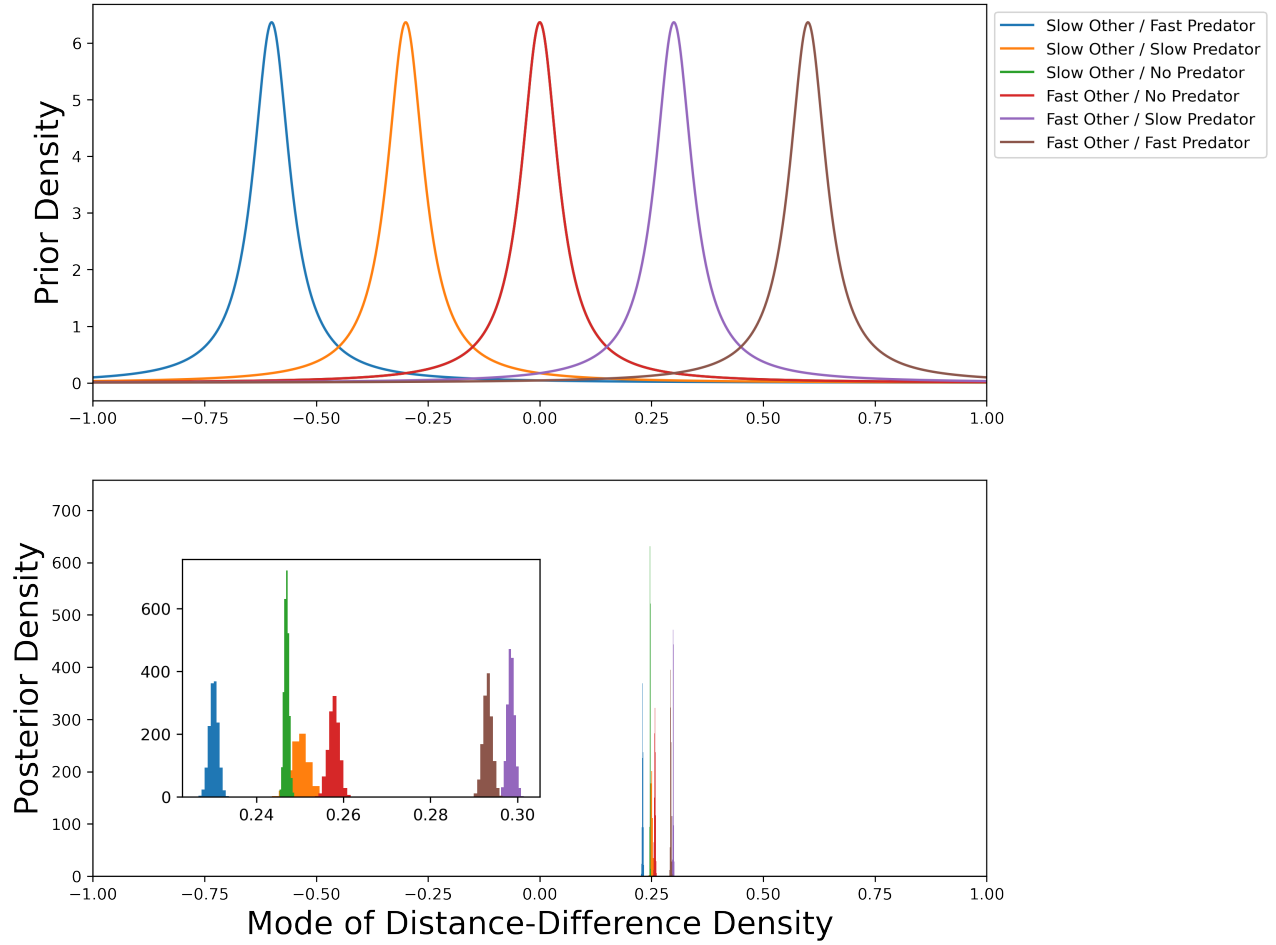


Figure 20: A visual comparison of the priors and posteriors of the modes ω_j of the distance-difference Beta likelihood. The inset plot is a zoom into the narrow range of the posterior distributions.

Conclusion

The vast majority of work investigating safety decisions and the intersectional effects of affect and social contexts on decision-making have focused on discrete decisions in ecologically unrealistic environments. The novelty of this work is that it builds the task on the basis of continuous decisions and naturalistic virtual ecologies. It is pertinent that any future work building off of this task environment or attempting to remedy faults in the design of this task environment remain true to those principles. Otherwise, we are doomed to a psychology whose models and explanations cannot be generalized to phenomena outside those generated artificially in the laboratory, a psychology not suited to our world.

Data availability

CSV files storing all data presented in this manuscript are available upon request.

Code availability

Code for the task web app and all analyses and modeling will be made available upon request.

Acknowledgements

The author would like to thank software engineer Paul Wheeler for his critical assistance in debugging the web app for the study.

<https://www.paulwheeler.us/resume.html>

References

- [1] Geoffrey Karl Aguirre, Marcelo Gomes Mattar, and Lucía Magis-Weinberg. “de Bruijn cycles for neural decoding”. In: *NeuroImage* 56.3 (June 2011), pp. 1293–1300. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2011.02.005. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1053811911001431> (visited on 01/31/2023).
- [2] Davide Albanese et al. “A practical tool for maximal information coefficient analysis”. In: *GigaScience* 7.4 (Apr. 1, 2018). ISSN: 2047-217X. DOI: 10.1093/gigascience/giy032. URL: <https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giy032/4958979> (visited on 01/31/2023).
- [3] Maribel Baldellou and S. Peter Henzi. “Vigilance, predator detection and the presence of supernumerary males in vervet monkey troops”. In: *Animal Behaviour* 43.3 (Mar. 1992), pp. 451–461. ISSN: 00033472. DOI: 10.1016/S0003-3472(05)80104-6. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0003347205801046> (visited on 01/18/2023).
- [4] Lane Beckes and James A. Coan. “Social Baseline Theory: The Role of Social Proximity in Emotion and Economy of Action: Social Baseline Theory”. In: *Social and Personality Psychology Compass* 5.12 (Dec. 2011), pp. 976–988. ISSN: 17519004. DOI: 10.1111/j.1751-9004.2011.00400.x. URL: <https://onlinelibrary.wiley.com/doi/10.1111/j.1751-9004.2011.00400.x> (visited on 01/18/2023).

- [5] D Caroline Blanchard et al. "Human defensive behaviors to threat scenarios show parallels to fear- and anxiety-related defense patterns of non-human mammals". In: *Neuroscience & Biobehavioral Reviews* 25.7 (Dec. 2001), pp. 761–770. ISSN: 01497634. DOI: 10.1016/S0149-7634(01)00056-2. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0149763401000562> (visited on 01/18/2023).
- [6] Bob Carpenter et al. "Stan: A Probabilistic Programming Language". In: *Journal of Statistical Software* 76.1 (2017). ISSN: 1548-7660. DOI: 10.18637/jss.v076.i01. URL: <http://www.jstatsoft.org/v76/i01/> (visited on 01/31/2023).
- [7] Charles S. Carver and Teri L. White. "Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales." In: *Journal of Personality and Social Psychology* 67.2 (Aug. 1994), pp. 319–333. ISSN: 1939-1315, 0022-3514. DOI: 10.1037/0022-3514.67.2.319. URL: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.67.2.319> (visited on 01/31/2023).
- [8] Xianwei Che et al. "A Systematic Review of the Processes Underlying the Main and the Buffering Effect of Social Support on the Experience of Pain". In: *The Clinical Journal of Pain* 34.11 (Nov. 2018), pp. 1061–1076. ISSN: 0749-8047. DOI: 10.1097/AJP.0000000000000624. URL: <https://journals.lww.com/00002508-201811000-00010> (visited on 01/18/2023).
- [9] John P. Christianson et al. "Safety Signals Mitigate the Consequences of Uncontrollable Stress Via a Circuit Involving the Sensory Insular Cortex and Bed Nucleus of the Stria Terminalis". In: *Biological Psychiatry* 70.5 (Sept. 2011), pp. 458–464. ISSN: 00063223. DOI: 10.1016/j.biopsych.2011.04.004. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0006322311003787> (visited on 01/18/2023).
- [10] James A. Coan, Shelley Kastle, et al. "Mutuality and the social regulation of neural threat responding". In: *Attachment & Human Development* 15.3 (May 2013), pp. 303–315. ISSN: 1461-6734, 1469-2988. DOI: 10.1080/14616734.2013.782656. URL: <http://www.tandfonline.com/doi/abs/10.1080/14616734.2013.782656> (visited on 01/18/2023).
- [11] James A. Coan, Hillary S. Schaefer, and Richard J. Davidson. "Lending a Hand: Social Regulation of the Neural Response to Threat". In: *Psychological Science* 17.12 (Dec. 2006), pp. 1032–1039. ISSN: 0956-7976, 1467-9280. DOI: 10.1111/j.1467-9280.2006.01832.x. URL: <http://journals.sagepub.com/doi/10.1111/j.1467-9280.2006.01832.x> (visited on 01/18/2023).

- [12] Mark H. Davis. "Measuring individual differences in empathy: Evidence for a multidimensional approach." In: *Journal of Personality and Social Psychology* 44.1 (Jan. 1983), pp. 113–126. ISSN: 1939-1315, 0022-3514. DOI: 10.1037/0022-3514.44.1.113. URL: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.44.1.113> (visited on 01/31/2023).
- [13] Ilan Eshel, Emilia Sansone, and Avner Shaked. "On the evolution of group-escape strategies of selfish prey". In: *Theoretical Population Biology* 80.2 (Sept. 2011), pp. 150–157. ISSN: 00405809. DOI: 10.1016/j.tpb.2011.06.005. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0040580911000578> (visited on 01/18/2023).
- [14] Michael S. Fanselow and Laurie S. Lester. "A functional behavioristic approach to aversively motivated behavior: Predatory imminence as a determinant of the topography of defensive behavior." In: 1988.
- [15] Charles W. Fox and Stephen J. Roberts. "A tutorial on variational Bayesian inference". In: *Artificial Intelligence Review* 38.2 (Aug. 2012), pp. 85–95. ISSN: 0269-2821, 1573-7462. DOI: 10.1007/s10462-011-9236-8. URL: <http://link.springer.com/10.1007/s10462-011-9236-8> (visited on 01/31/2023).
- [16] Andrew Gelman. *Bayesian data analysis*. Third edition. Chapman & Hall/CRC texts in statistical science. Boca Raton: CRC Press, 2014. 661 pp. ISBN: 9781439840955.
- [17] Jesse Graham and Jonathan Haidt. "Sacred values and evil adversaries: A moral foundations approach." In: *The social psychology of morality: Exploring the causes of good and evil*. Ed. by Mario Mikulincer and Phillip R. Shaver. Washington: American Psychological Association, 2012, pp. 11–31. ISBN: 9781433810114. DOI: 10.1037/13091-001. URL: <http://content.apa.org/books/13091-001> (visited on 01/31/2023).
- [18] W.D. Hamilton. "Geometry for the selfish herd". In: *Journal of Theoretical Biology* 31.2 (May 1971), pp. 295–311. ISSN: 00225193. DOI: 10.1016/0022-5193(71)90189-5. URL: <https://linkinghub.elsevier.com/retrieve/pii/0022519371901895> (visited on 01/18/2023).
- [19] Erica A. Hornstein and Naomi I. Eisenberger. "Unpacking the buffering effect of social support figures: Social support attenuates fear acquisition". In: *PLOS ONE* 12.5 (May 2, 2017). Ed. by Kirk Warren Brown, e0175891. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0175891. URL: <https://dx.plos.org/10.1371/journal.pone.0175891> (visited on 01/18/2023).

- [20] Erica A. Hornstein, Michael S. Fanselow, and Naomi I. Eisenberger. "A Safe Haven: Investigating Social-Support Figures as Prepared Safety Stimuli". In: *Psychological Science* 27.8 (Aug. 2016), pp. 1051–1060. ISSN: 0956-7976, 1467-9280. DOI: 10.1177/0956797616646580. URL: <http://journals.sagepub.com/doi/10.1177/0956797616646580> (visited on 01/18/2023).
- [21] Alp Kucukelbir et al. *Automatic Variational Inference in Stan*. June 12, 2015. DOI: 10.48550/arXiv.1506.03431. arXiv: 1506.03431[stat]. URL: <http://arxiv.org/abs/1506.03431> (visited on 01/31/2023).
- [22] P.F. Lovibond and S.H. Lovibond. "The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories". In: *Behaviour Research and Therapy* 33.3 (Mar. 1995), pp. 335–343. ISSN: 00057967. DOI: 10.1016/0005-7967(94)00075-U. URL: <https://linkinghub.elsevier.com/retrieve/pii/000579679400075U> (visited on 01/31/2023).
- [23] Michael Mendl and Elizabeth S. Paul. "Animal affect and decision-making". In: *Neuroscience & Biobehavioral Reviews* 112 (May 2020), pp. 144–163. ISSN: 01497634. DOI: 10.1016/j.neubiorev.2020.01.025. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0149763419305202> (visited on 01/18/2023).
- [24] Dean Mobbs. "The ethological deconstruction of fear(s)". In: *Current Opinion in Behavioral Sciences* 24 (Dec. 2018), pp. 32–37. ISSN: 23521546. DOI: 10.1016/j.cobeha.2018.02.008. URL: <https://linkinghub.elsevier.com/retrieve/pii/S235215461730267X> (visited on 01/18/2023).
- [25] Dean Mobbs, Cindy C. Hagan, et al. "The ecology of human fear: survival optimization and the nervous system". In: *Frontiers in Neuroscience* 9 (Mar. 18, 2015). ISSN: 1662-453X. DOI: 10.3389/fnins.2015.00055. URL: http://www.frontiersin.org/Evolutionary_Psychology_and_Neuroscience/10.3389/fnins.2015.00055/abstract (visited on 01/18/2023).
- [26] Dean Mobbs, Toby Wise, et al. "Promises and challenges of human computational ethology". In: *Neuron* 109.14 (July 2021), pp. 2224–2238. ISSN: 08966273. DOI: 10.1016/j.neuron.2021.05.021. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0896627321003743> (visited on 01/21/2023).
- [27] L. J. Morrell, G. D. Ruxton, and R. James. "Spatial positioning in the selfish herd". In: *Behavioral Ecology* 22.1 (Jan. 1, 2011), pp. 16–22. ISSN: 1045-2249, 1465-7279. DOI: 10.1093/beheco/

- arq157. URL: <https://academic.oup.com/beheco/article-lookup/doi/10.1093/beheco/arq157> (visited on 01/18/2023).
- [28] Thomas L. Morton et al. "The Selfish Herd Revisited: Do Simple Movement Rules Reduce Relative Predation Risk?" In: *Journal of Theoretical Biology* 167.1 (Mar. 1994), pp. 73–79. ISSN: 00225193. DOI: 10.1006/jtbi.1994.1051. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0022519384710514> (visited on 01/18/2023).
- [29] James E. Orpwood et al. "Minnows and the selfish herd: effects of predation risk on shoaling behaviour are dependent on habitat complexity". In: *Animal Behaviour* 76.1 (July 2008), pp. 143–152. ISSN: 00033472. DOI: 10.1016/j.anbehav.2008.01.016. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0003347208001292> (visited on 01/18/2023).
- [30] Yafeng Pan, Andreas Olsson, and Armita Golkar. "Social safety learning: Shared safety abolishes the recovery of learned threat". In: *Behaviour Research and Therapy* 135 (Dec. 2020), p. 103733. ISSN: 00057967. DOI: 10.1016/j.brat.2020.103733. URL: <https://linkinghub.elsevier.com/retrieve/pii/S000579672030187X> (visited on 01/18/2023).
- [31] Song Qi et al. "A Collaborator's Reputation Can Bias Decisions and Anxiety under Uncertainty". In: *The Journal of Neuroscience* 38.9 (Feb. 28, 2018), pp. 2262–2269. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.2337-17.2018. URL: <https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.2337-17.2018> (visited on 01/18/2023).
- [32] David N. Reshef et al. "Detecting Novel Associations in Large Data Sets". In: *Science* 334.6062 (Dec. 16, 2011), pp. 1518–1524. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1205438. URL: <https://www.science.org/doi/10.1126/science.1205438> (visited on 01/31/2023).
- [33] Yakir A. Reshef et al. *Theoretical Foundations of Equitability and the Maximal Information Coefficient*. May 12, 2015. DOI: 10.48550/arXiv.1408.4908. arXiv: 1408.4908[cs, math, q-bio, stat]. URL: <http://arxiv.org/abs/1408.4908> (visited on 01/31/2023).
- [34] Craig W. Reynolds. "Flocks, herds and schools: A distributed behavioral model". In: *Proceedings of the 14th annual conference on Computer graphics and interactive techniques - SIGGRAPH '87*. the 14th annual conference. Not Known: ACM Press, 1987, pp. 25–34. ISBN: 9780897912273. DOI: 10.1145/37401.37406.

- URL: <http://portal.acm.org/citation.cfm?doid=37401.37406> (visited on 01/31/2023).
- [35] Filipe Cristovão Ribeiro da Cunha. "Sex, cooperation, and deception: anti-predatory behavior beyond avoiding death". PhD thesis. University of Zurich, 2017. DOI: 10.5167/UZH-146575. URL: <https://www.zora.uzh.ch/id/eprint/146575> (visited on 01/18/2023).
- [36] Carel P. Schaik et al. "Male anti-predation services in primates as costly signalling? A comparative analysis and review". In: *Ethology* 128.1 (Jan. 2022), pp. 1–14. ISSN: 0179-1613, 1439-0310. DOI: 10.1111/eth.13233. URL: <https://onlinelibrary.wiley.com/doi/10.1111/eth.13233> (visited on 01/18/2023).
- [37] Carel P. van Schaik and Mark Hörstermann. "Predation risk and the number of adult males in a primate group: a comparative test". In: *Behavioral Ecology and Sociobiology* 35.4 (Oct. 1994), pp. 261–272. ISSN: 0340-5443, 1432-0762. DOI: 10.1007/BF00170707. URL: <http://link.springer.com/10.1007/BF00170707> (visited on 01/18/2023).
- [38] Craig Stanford. "Predation and Male Bonds in Primate Societies". In: *Behaviour* 135.4 (1998), pp. 513–533. ISSN: 0005-7959, 1568-539X. DOI: 10.1163/156853998793066212. URL: https://brill.com/view/journals/beh/135/4/article-p513_6.xml (visited on 01/18/2023).
- [39] Craig B. Stanford. "Avoiding Predators: Expectations and Evidence in Primate Antipredator Behavior". In: *International Journal of Primatology* 23.4 (2002), pp. 741–757. ISSN: 01640291. DOI: 10.1023/A:1015572814388. URL: <http://link.springer.com/10.1023/A:1015572814388> (visited on 01/18/2023).
- [40] Ellen Tedeschi et al. "Inferences of Others' Competence Reduces Anticipation of Pain When under Threat". In: *Journal of Cognitive Neuroscience* 27.10 (Oct. 1, 2015), pp. 2071–2078. ISSN: 0898-929X, 1530-8898. DOI: 10.1162/jocn_a_00843. URL: <https://direct.mit.edu/jocn/article/27/10/2071/28381/Inferences-of-0thers-Competence-Reduces> (visited on 01/18/2023).
- [41] Steven V. Viscido, Matthew Miller, and David S. Wethey. "The Dilemma of the Selfish Herd: The Search for a Realistic Movement Rule". In: *Journal of Theoretical Biology* 217.2 (July 2002), pp. 183–194. ISSN: 00225193. DOI: 10.1006/jtbi.2002.3025. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0022519302930250> (visited on 01/18/2023).

- [42] Steven V. Viscido and David S. Wetthey. "Quantitative analysis of fiddler crab flock movement: evidence for 'selfish herd' behaviour". In: *Animal Behaviour* 63.4 (Apr. 2002), pp. 735–741. ISSN: 00033472. DOI: 10.1006/anbe.2001.1935. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0003347201919359> (visited on 01/18/2023).
- [43] Seng Bum Michael Yoo, Benjamin Yost Hayden, and John M. Pearson. "Continuous decisions". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 376.1819 (Mar. 2021), p. 20190664. ISSN: 0962-8436, 1471-2970. DOI: 10.1098/rstb.2019.0664. URL: <https://royalsocietypublishing.org/doi/10.1098/rstb.2019.0664> (visited on 01/21/2023).
- [44] Jiajin Yuan et al. "Social Company Disrupts Fear Memory Renewal: Evidence From Two Rodent Studies". In: *Frontiers in Neuroscience* 12 (Aug. 17, 2018), p. 565. ISSN: 1662-453X. DOI: 10.3389/fnins.2018.00565. URL: <https://www.frontiersin.org/article/10.3389/fnins.2018.00565/full> (visited on 01/18/2023).