Monograph Proposal: A Computational Framework for the Analysis and Mitigation of Cognitive Biases in Human Decision-Making Processes

Prospective Publication Venue: Cognitive Systems Journal, or an alternative peer-reviewed academic repository of comparable standing, such as Topics in Cognitive Science or Computational Brain & Behavior.
Anticipated Chronological Framework for Completion (Consequent to Revised Analytical Schema): A period projected at twelve to fourteen calendar months is allocated for the consummation of this research program. This timeline incorporates dedicated phases for supplementary validation studies, particularly concerning the efficacy and robustness of intervention strategies, and for the meticulous documentation of unsuccessful or sub-optimal methodological iterations, which are considered integral to the scientific contribution.

I. Conceptual Abstract (Derived from Reconstructed Abstract & Subsequent Revisions)

This manuscript is intended to introduce, formally specify, and empirically validate a comprehensive computational framework designed for the systematic modeling, simulation-based analysis, and potential mitigation of the impact exerted by common cognitive biases—including, but not limited to, confirmation bias, anchoring effects, and the availability heuristic—upon human decision-making processes across diverse contexts. The core of this research will detail the architectural design and algorithmic implementation of an agent-based model, wherein individual agents are endowed with parameterized mechanisms intended to replicate these specified cognitive biases. The veridicality of this model will be rigorously assessed through its validation against empirical data derived from meticulously designed human subject experiments, which are specifically constructed to elicit and quantify the targeted biases. The empirical investigation will aim to demonstrate the computational model's capacity to replicate known patterns of biased decision-making with a pre-specified level of accuracy, substantiated by appropriate statistical confidence intervals. Furthermore, this research will explore the efficacy of computationally-derived intervention strategies, designed within the modeling framework to counteract or ameliorate the influence of these biases. The effectiveness of such interventions will be reported with realistic effect sizes, conservative statistical estimates, and corresponding confidence intervals, thereby contributing to both the theoretical understanding of cognitive biases and the development of evidence-based debiasing techniques.

II. Research Components & Methodological Protocol

A. Computational Model Development: Architecture and Formalization

1. **Bias Formalization: Mathematical and Algorithmic Specification:**
    - **Confirmation Bias:** This bias, characterized by the propensity to seek, interpret, favor, and recall information in a manner that confirms or supports pre-existing beliefs or values, will be modeled as a systematic modification of Bayesian belief updating mechanisms. For instance, the updating rule may be represented as $P_{biased}(H|E) = P(H|E)^{\beta} / [P(H|E)^{\beta} + (1-P(H|E))^{\beta}]$, where $P(H|E)$ is the normative posterior probability of hypothesis H given evidence E, and $\beta > 1$ is a parameter quantifying the strength of the confirmatory tendency. The estimation or systematic variation of β across simulated agents or conditions will be a key component of the model. Further algorithmic instantiations may involve biased information search strategies, wherein agents preferentially sample evidence congruent with their current hypotheses, or differential weighting of belief-consistent versus belief-inconsistent evidence during integration.
    - **Anchoring Bias:** This cognitive heuristic, describing the common human tendency to rely too heavily on an initial piece of information (the "anchor") offered when making decisions, will be modeled by representing the agent's final estimate as a weighted average: $\text{Estimate} = \alpha \times \text{Anchor} + (1-\alpha) \times \text{Normative\_Integration\_of\_Evidence} + \epsilon$. Here, $\alpha \in [0,1]$ represents the anchoring strength or the weight assigned to the initial anchor value, $\text{Normative\_Integration\_of\_Evidence}$ represents an unbiased estimate derived from all available information subsequent to the anchor, and ϵ denotes a stochastic noise term. The sources of anchors (e.g., externally provided, self-generated) and the mechanisms by which they influence subsequent information processing will be explicitly modeled.
    - **Availability Heuristic:** This mental shortcut, which relies on immediate examples that come to a given person's mind when evaluating a specific topic, concept, method or decision, will be modeled by formalizing the perceived probability or frequency of an event as a function of its cognitive accessibility. This may be represented as $P_{\text{perceived}}(\text{event}) = f(\text{recency, vividness, frequency\_of\_retrieval})$. The function f will map these factors, potentially operationalized through simulated memory trace strengths or retrieval dynamics (e.g., using activation thresholds or decay functions like $A_i(t) = B_i + \sum_{j=1}^{n} M_{ji} \exp(-d_{ji} \cdot t)$, where $A_i(t)$ is activation of item i at time t), to a subjective probability estimate.
    - **Modeling of Bias Interactions:** The framework will incorporate mechanisms for modeling the potential interactions between these different cognitive biases, as human decision-making is rarely influenced by a single bias in isolation. This may

involve exploring sequential effects (e.g., an initial anchor influencing the subsequent search for confirmatory evidence), additive or multiplicative effects on decision thresholds or evidence accumulation parameters, or more complex, non-linear interactions mediated by shared underlying cognitive resources (e.g., working memory capacity).

2. **Agent-Based Simulation Framework: Design and Calibration:**
    - The agent architecture will be designed with distinct cognitive components, including perceptual modules for processing environmental stimuli, a parameterized memory system (potentially with separate short-term and long-term stores and specific encoding/retrieval dynamics), learning rules (e.g., reinforcement learning, Bayesian updating), and decision-making modules that incorporate the formalized bias mechanisms. Each agent will possess a set of parameters defining its susceptibility to different biases (e.g., individual β values for confirmation bias, α values for anchoring).
    - A comprehensive suite of decision-making scenarios will be created, designed to elicit the targeted cognitive biases under controlled conditions. These scenarios will span various domains, such as financial investment decisions (e.g., evaluating stocks based on initial price movements), medical diagnostic judgments (e.g., assessing patient symptoms given preliminary hypotheses), social judgments (e.g., forming impressions based on limited information), and consumer choice paradigms. The scenarios will allow for systematic manipulation of information presentation, anchor values, and evidence availability.
    - Mechanisms for calibrating the model's free parameters, particularly those governing bias susceptibility and cognitive capacities, will be implemented. This may involve employing techniques such as grid search over parameter spaces, more sophisticated optimization algorithms (e.g., genetic algorithms, particle swarm optimization), or Bayesian estimation methods (e.g., Markov Chain Monte Carlo (MCMC) sampling) to find parameter settings that best reproduce aggregate human behavioral patterns or individual participant data.

B. Empirical Validation with Human Subjects: Protocol and Data Analysis

1. **Human Studies Design and Execution (Contingent upon Institutional Review Board Sanction):**
    - A series of controlled laboratory experiments will be designed to elicit and measure specific cognitive biases in human participants. These experiments will employ established paradigms from the cognitive psychology literature, such as information search tasks (e.g., Wason selection task variants for confirmation bias), numerical estimation tasks with manipulated anchors, and frequency judgment or recall tasks designed to probe the availability heuristic. The experimental design will allow for precise manipulation of independent variables hypothesized to influence the magnitude of these biases.

- Participants, numbering approximately n≈200–250 per study (with the precise sample size determined by a formal power analysis incorporating an uncertainty factor: $n=(Z\alpha/2+Z\beta)2\times2\sigma2/\delta2\times(1+\epsilon)$), will be recruited from university populations or online participant pools (e.g., Prolific, MTurk), with appropriate demographic information collected. Ethical considerations, including informed consent, minimization of psychological discomfort, and thorough debriefing, will be paramount.
- A rich set of behavioral data will be collected, including participants' overt choices, confidence ratings associated with their decisions, response latencies, and potentially process-tracing measures such as eye-tracking data or information acquisition sequences, where feasible and informative.

2. **Model-Data Comparison and Iterative Refinement:**
   - The outputs generated by the agent-based simulations (e.g., distributions of choices, mean error rates, patterns of confidence judgments) will be systematically compared with the corresponding human behavioral data obtained from the empirical studies. This comparison will extend beyond aggregate statistics to include, where possible, the analysis of error patterns, response time distributions, and correlations between model parameters and individual differences in human performance.
   - A variety of statistical methods will be employed to quantify the goodness-of-fit between the model and the data. These may include Chi-square tests for categorical data, Kullback-Leibler (KL) divergence for comparing probability distributions, root mean squared error (RMSE) for continuous predictions, and Bayesian model comparison techniques (e.g., Bayes factors, Deviance Information Criterion (DIC)) for selecting among competing model variants or parameterizations.
   - An iterative process of model refinement will be undertaken. Discrepancies between model predictions and empirical observations will inform targeted modifications to the model's parameters, architectural structure, or the formalization of its bias mechanisms, with the goal of progressively enhancing the model's descriptive adequacy and predictive power.

C. Intervention Strategy Development & Empirical Testing

1. **Computationally-Derived Debiasing Interventions:**
   - Based on the insights derived from the computational model (e.g., identifying specific information processing bottlenecks or parameter sensitivities that exacerbate biases), a range of debiasing strategies will be designed. These may include, for example: (i) altering the mode or sequence of information presentation to counteract confirmation tendencies (e.g., forcing consideration of disconfirming evidence first); (ii) providing explicitly generated counterfactual scenarios or alternative anchors to mitigate anchoring effects; (iii) utilizing structured reflection prompts or decision aids (e.g., checklists, pre-mortem analyses) to encourage more systematic and less heuristic-driven processing; (iv)

training interventions aimed at increasing metacognitive awareness of biases.
    - The design of these interventions will be guided by the principle of identifying leverage points within the modeled cognitive architecture where targeted modifications can yield maximal bias reduction with minimal disruption to otherwise adaptive cognitive functions.
2. **Intervention Testing: Simulation and Human Subject Experiments:**
    - The efficacy of the designed interventions will first be evaluated within the agent-based simulation framework. This will involve comparing the decision outcomes of simulated agents operating with and without the interventions across various scenarios. This in-silico testing allows for rapid iteration and refinement of intervention designs.
    - Promising interventions, identified through simulation, will subsequently be tested, where feasible and ethically appropriate, in new human subject studies. These studies will typically employ a between-subjects or within-subjects experimental design, comparing decision-making performance in the presence of the intervention against a control condition (e.g., no intervention, or a placebo intervention).

D. Comprehensive Statistical Analysis Plan

- **Model Fitting and Parameter Estimation:** Advanced statistical techniques will be utilized for model fitting and parameter estimation, such as maximum likelihood estimation (MLE), Bayesian estimation (e.g., MCMC), or hierarchical Bayesian modeling to account for individual differences. Model selection criteria (e.g., Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), DIC) will be employed to compare the relative parsimony and fit of different model variants.
- **Comparison of Model Predictions to Human Data:** Beyond simple mean comparisons, distributional analyses (e.g., comparing histograms or cumulative distribution functions of choices or response times) and correlational analyses (e.g., correlating model-derived bias susceptibility parameters with individual participants' observed bias magnitudes) will be conducted. Effect sizes (e.g., Cohen's d, R2) and their confidence intervals will be reported for all key comparisons.
- **Evaluation of Intervention Effectiveness:** The statistical significance and practical magnitude of intervention effects will be assessed using appropriate inferential tests (e.g., t-tests, ANOVA, ANCOVA, mixed-effects models to account for repeated measures or nested data structures). Effect sizes for intervention efficacy (e.g., reduction in bias score, improvement in decision accuracy) will be reported with 95% confidence intervals.
- **Correction for Multiple Comparisons:** Where multiple hypotheses are tested or multiple outcome measures are analyzed, appropriate corrections for multiple comparisons (e.g., Bonferroni, Holm-Bonferroni, Benjamini-Hochberg FDR) will be applied to control the Type I error rate.
- **Robustness Checks and Sensitivity Analyses:** The robustness of the findings will be assessed through various checks, including sensitivity of model parameters to

variations in input data or minor changes in model assumptions, cross-validation procedures to evaluate out-of-sample predictive performance, and potentially simulation-based recovery studies to ensure that model parameters are identifiable from data of the type being collected.

III. Expected Outcomes & Performance Metrics (Revised & Verisimilar Projections with Enhanced Detail)

A. Model Accuracy in Replicating Observed Bias Patterns

- **Primary Point Estimate:** An accuracy of 86 (95% Confidence Interval: [82%, 90%]) is projected in the computational model's ability to predict or replicate human participants' biased choice patterns, when compared against chance-level performance or a null model that does not incorporate bias mechanisms. This accuracy metric will be operationalized, for example, as the percentage of experimental trials where the model correctly predicts the direction or magnitude of deviation from normative decision-making exhibited by human subjects.
- **Conservative Estimate (Lower Bound of Confidence Interval):** An 82% accuracy in replicating biased patterns is considered the conservative projection, which would still represent a substantial advancement in the computational modeling of these pervasive cognitive phenomena.
- **Elaboration and Nuance:** It is explicitly noted that the aggregate accuracy in replicating bias patterns will be further disaggregated and reported on a per-bias basis. Per-bias accuracy is anticipated to exhibit variability, reflecting the differential complexity and subtlety of the underlying cognitive mechanisms (e.g., Confirmation Bias: projected accuracy of 83% ± 5%; Anchoring Effects: projected accuracy of 88% ± 4%; Availability Heuristic: projected accuracy of 80% ± 6%). This differentiated reporting will provide more granular insights into the model's strengths and weaknesses. The implications of achieving such accuracy levels include enhanced theoretical understanding of how biases arise from specific computational mechanisms and the potential for using the model as a testbed for novel debiasing strategies.

B. Efficacy of Bias Reduction Interventions (Derived from Computational Modeling)

- **Primary Point Estimate (Averaged across tested interventions and targeted biases):** A mean reduction of 14 (95% Confidence Interval: [8%, 20%]) in the measured magnitude or observed frequency of biased decisions is projected as a result of applying the computationally-derived intervention strategies. This reduction will be quantified relative to control conditions where no intervention is applied.
- **Conservative Estimate (Lower Bound of Confidence Interval):** An 8% average reduction in bias manifestation is considered the conservative projection, which, if

achieved across multiple bias types, would nonetheless represent a practically significant outcome for improving decision quality.

- **Elaboration and Nuance:** The effectiveness of interventions is explicitly anticipated to be highly variable, contingent upon both the specific type of cognitive bias being targeted and the nature of the intervention strategy employed. Illustrative projections include:
    - Anchoring effects, when addressed via a "consider-the-opposite" cognitive strategy or by providing multiple, diverse anchors, might exhibit a reduction of 18.
    - Confirmation bias, when counteracted by interventions promoting a structured search for and evaluation of disconfirming evidence, might show a reduction of 11.
    - The availability heuristic, when mitigated by providing base-rate statistical information or prompting consideration of less salient but equally probable alternatives, might demonstrate a reduction of 13.
      The practical significance of a 14% average reduction lies in its potential to improve decision outcomes in domains where biases can have substantial negative consequences (e.g., financial planning, medical judgment, legal proceedings).

IV. Reporting of Failed Intervention Strategies, Null Results, & Iterative Model Refinements

- **Significance and Rationale for Comprehensive Reporting:** The transparent and detailed reporting of intervention strategies that proved ineffective or yielded null results, alongside the documentation of iterative refinements to the computational model, is considered essential for building a cumulative and robust science of debiasing. Such reporting prevents the "file drawer problem," contributes to a more accurate understanding of the complexities involved in modifying entrenched cognitive patterns, and provides valuable negative evidence that can guide future research efforts.
- **Exemplars for Comprehensive Documentation, Analysis, and Interpretation:**
    1. **Direct Bias Notification (Hypothesized Failure or Limited Efficacy with Adverse Consequences):**
        - Attempted Intervention: Simply informing simulated agents or human participants that they might be susceptible to a particular bias before or during a decision-making task.
        - Anticipated Outcome and Analysis: A minimal, potentially statistically insignificant, reduction in the manifestation of the targeted bias (e.g., a 2% ± 1.5% reduction) is anticipated. Critically, this intervention might also lead to unintended adverse consequences, such as a significant increase in decision time (e.g., a 40% ± 10% increase) due to induced meta-awareness or self-monitoring that does not translate into effective corrective action, or even a "rebound" effect where awareness paradoxically increases reliance on the bias. The analysis would explore the cognitive mechanisms

underlying such limited efficacy, potentially relating to dual-task interference or insufficient procedural knowledge for bias correction.

2. **Forced Perspective-Taking or "Consider the Alternative" Prompts (Hypothesized Mixed Results or User Resistance):**
    - Attempted Intervention: Requiring simulated agents or human participants to explicitly list reasons supporting an alternative choice or to articulate the perspective of an individual holding an opposing viewpoint.
    - Anticipated Outcome and Analysis: A modest reduction in certain biases (e.g., a 5% ± 3% reduction in confirmation bias) might be observed. However, this could be accompanied by reports of high user frustration, increased perceived cognitive effort, or even reactance, particularly if the intervention is perceived as overly prescriptive or time-consuming, thereby impacting its ecological validity and potential for adoption in real-world settings. The analysis would investigate the trade-off between objective bias reduction and subjective user experience.
3. **Automated Algorithmic Bias Correction or Nudging (Hypothesized Failure in User Acceptance or Trust):**
    - Attempted Intervention: A system that algorithmically detects a likely bias in the user's input or preliminary decision and automatically adjusts that input, suggests an alternative, or "nudges" the user towards a less biased option.
    - Anticipated Outcome and Analysis: While such a system might demonstrate high efficacy in pure simulation (i.e., when agents passively accept corrections), it is hypothesized to encounter low user acceptance (e.g., a greater than 70% rejection or manual override rate in human-in-the-loop experimental tests). This resistance may stem from a perceived lack of transparency in the correction mechanism, a feeling of reduced autonomy or agency on the part of the user, or insufficient trust in the algorithmic adjustment. The study would explore factors influencing trust and acceptance of such automated interventions.

- The process of documenting iterative refinements to the computational model itself will be detailed, highlighting specific changes made to its architecture, parameterization, or algorithmic components based on discrepancies observed during the comparison with initial empirical data or based on the failure of early model versions to adequately capture specific biased phenomena. This may involve, for example, version control logs for the model code, accompanied by a narrative explaining the rationale for each significant modification and its impact on model performance and plausibility.

V. Acknowledgment of Inherent Methodological & Conceptual Trade-offs, Limitations, and Boundary Conditions

- **Intervention Effectiveness versus User Acceptance, Cognitive Effort, and**

Perceived Autonomy:

- A formal mathematical model, such as User Acceptance $=k1\times e-k2\times InterventionIntensity-k3\times PerceivedEffort$, may be posited and explored, where Intervention Intensity could be a composite measure of the intervention's prescriptiveness, intrusiveness, or time requirement.
- A detailed narrative discussion will acknowledge that more intensive or cognitively demanding interventions, while potentially more effective in objectively reducing bias magnitude, might be less likely to be adopted or consistently utilized by individuals in real-world settings due to increased cognitive load, time constraints, or a desire to maintain decisional autonomy. The "cost" of debiasing, in terms of effort and acceptance, is a critical factor. The implications of these trade-offs for the design of practical decision support systems will be thoroughly examined.

- **Bias Reduction versus Decision Speed, Efficiency, and Potential for Over-Correction:**
  - A mathematical model, such as $\Delta DecisionTime=k3\times(BiasReduction)k4+k5\times TaskComplexity$, may be explored to quantify the relationship between the extent of bias reduction achieved and the concomitant increase in decision latency, potentially moderated by task complexity.
  - The narrative will elaborate on the observation that debiasing often necessitates more deliberate, systematic (System 2) processing, which can inherently slow down decision-making compared to faster, heuristic-driven (System 1) processing. The potential for interventions to induce over-correction, leading to new forms of error or an excessive and inefficient allocation of cognitive resources to relatively minor decisions, will also be considered as a critical limitation.
- **Model Generalizability and Representational Fidelity:**
  - A frank discussion will address the inherent limitations of the computational model in capturing the full spectrum and nuance of human cognitive biases, which are influenced by a vast array of individual differences (e.g., cognitive styles, personality traits, domain expertise), emotional states, and socio-cultural contexts that may not be fully represented within the model's architecture or parameter space. The applicability of the model's findings across different populations and decision domains will be carefully circumscribed.
  - The challenges of achieving high representational fidelity—that is, ensuring that the model's internal mechanisms not only reproduce behavioral outcomes but also genuinely reflect the underlying cognitive processes—will be acknowledged. The model should be viewed as a formal instantiation of a particular theory of biased cognition, subject to ongoing refinement and testing against alternative theoretical accounts.
- **Ecological Validity and the Laboratory-to-Real-World Gap:**
  - The inherent differences between controlled laboratory tasks, often designed for experimental tractability, and complex, ill-defined, high-stakes decision-making in

real-world environments will be explicitly acknowledged. Factors such as dynamic information environments, severe time pressure, significant personal or organizational consequences, and the influence of social dynamics, which are often absent or simplified in laboratory settings, can profoundly modulate the expression and impact of cognitive biases. The implications of these differences for the generalizability of the study's findings and the practical applicability of the proposed interventions will be critically assessed.

VI. Proposed Structure of the Monograph or Principal Publication (Adhering to Disciplinary Norms and Journal-Specific Guidelines)

1. **Abstract:** A succinct and informative summary (typically constrained to 200-300 words) encapsulating the principal research objectives, the core methodological innovations (i.e., the development of the computational framework and intervention strategies), the most salient empirical findings concerning model accuracy and intervention efficacy (reported with confidence intervals), and a clear articulation of the study's primary contributions to the scientific understanding of cognitive biases and their mitigation.
2. **Introduction:** A comprehensive exposition of the theoretical and practical importance of understanding and mitigating cognitive biases in human decision-making. This section will clearly define the scope of the research, articulate the specific research questions and hypotheses, provide an overview of the computational modeling approach adopted, and briefly outline the structure of the ensuing manuscript.
3. **Theoretical Background and Literature Review:** A thorough review of the relevant psychological literature on the targeted cognitive biases (confirmation, anchoring, availability), including their operational definitions, documented empirical manifestations, putative cognitive and, where known, neural underpinnings, and existing theoretical accounts. This section will also survey prior computational modeling efforts in this domain, highlighting their strengths, limitations, and the specific gaps addressed by the current research.
4. **The Computational Model: Architecture, Formalization, and Implementation:** A detailed and technically precise description of the agent-based model's architecture, including the mathematical formalizations of the bias mechanisms, the specification of agent parameters, the design of the simulation environment, and the technical details of its software implementation. Sufficient detail will be provided to permit independent replication and extension of the model.
5. **Empirical Validation: Human Subject Studies and Model-Data Comparison:** A comprehensive account of the methods employed in the human subject experiments, including participant recruitment and characteristics, experimental designs, task procedures, and data collection protocols. This will be followed by a detailed presentation of the results comparing the computational model's predictions to the

observed human behavioral data, including quantitative measures of model fit and discrepancy analyses.

6. **Development and Evaluation of Intervention Strategies:** A clear description of the design rationale for the computationally-derived debiasing interventions. This section will present the results from both in-silico simulation-based testing of these interventions and, where applicable, from empirical testing with human participants, focusing on their efficacy in reducing bias and their impact on decision quality.
7. **General Discussion:** A thorough interpretation of the overall pattern of findings, situating them within the broader theoretical context. This section will include a critical evaluation of the model's limitations and boundary conditions, a nuanced discussion of the effectiveness and practical applicability of the tested interventions, an explicit consideration of the identified trade-offs, and an articulation of the principal theoretical and practical implications of the research for decision support system design, educational initiatives, and public policy. Directions for future research will also be proposed.
8. **Conclusion:** A concise recapitulation of the main research contributions, the key empirical findings, and their overarching significance for the fields of cognitive science, artificial intelligence, and decision theory.
9. **Supplementary Materials:** Provision of extensive supplementary information, potentially including the full source code for the computational model and simulation environment, detailed protocols for the human subject studies, complete statistical analysis scripts, extended tables of results, additional figures, and in-depth analyses of failed intervention strategies or model iterations that could not be accommodated within the main body of the manuscript due to space constraints.

VII. Ethical Considerations & Adherence to Open Science Principles and Best Practices

- **Institutional Review Board (IRB) Approval and Ethical Conduct of Research:** For all human subject studies conducted as part of this research program, prior approval will be sought and obtained from the relevant Institutional Review Board or ethics committee. All research activities will be conducted in strict adherence to the approved protocols and the ethical principles outlined in the Declaration of Helsinki and relevant national and institutional guidelines.
- **Participant Well-being, Informed Consent, and Debriefing:** Paramount importance will be accorded to ensuring the psychological and emotional well-being of all research participants. Decision-making tasks will be designed to minimize the potential for undue stress or psychological discomfort. A comprehensive informed consent process will be implemented, ensuring that participants are fully aware of the nature of the study, any potential risks and benefits, their rights to confidentiality and withdrawal, and how their data will be used, prior to their agreement to participate. Following participation, a thorough debriefing procedure will be employed to explain the research aims and, where appropriate, to provide educational information regarding cognitive biases in a

manner that avoids inducing undue self-criticism or distress.

- **Transparency in Reporting and Methodological Rigor:** A commitment to complete transparency in the reporting of research methods and findings will be maintained. All model assumptions, parameter choices, analytical decisions, and limitations of the research will be clearly and explicitly stated. Both statistically significant and non-significant findings, as well as results from unsuccessful or exploratory analyses, will be reported to provide a balanced and comprehensive account of the research.
- **Adherence to Open Science Principles and Data/Code Sharing:** In alignment with contemporary best practices in open science, a steadfast commitment will be made to pre-registering the hypotheses, study designs, and primary analysis plans for all human subject studies on a recognized public platform (e.g., Open Science Framework, AsPredicted.org). Furthermore, efforts will be made to make the computational model code, simulation environments, analytical scripts, and appropriately anonymized and de-identified empirical datasets publicly available through reputable repositories (e.g., GitHub, Zenodo, OSF), under permissive licenses, to the fullest extent ethically and legally permissible. This practice is intended to facilitate reproducibility, encourage secondary data analysis, and foster collaborative scientific advancement within the research community, adhering to the FAIR (Findable, Accessible, Interoperable, Reusable) data principles.

This comprehensive and expanded outline for Monograph 3 endeavors to provide a robust and scientifically rigorous framework for the investigation of cognitive biases through computational modeling, emphasizing meticulous empirical validation, the realistic assessment of intervention strategies, and a transparent, self-critical approach to the reporting of the research process and its outcomes.