Beta Oscillations Mediate Responses to Counterfactual Feedback During Decision-Making

Alexandra Fink (alexandra.fink@icahn.mssm.edu)
Nash Family Department of Neuroscience and the Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY

Salman Qasim (salman.qasim@mssm.edu)
The Center for Computational Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY
The Nash Family Center for Advanced Circuit Therapeutics, Icahn School of Medicine at Mount Sinai, New York, NY

Jacqueline Overton (jacqueline.overton@mssm.edu)
The Nash Family Center for Advanced Circuit Therapeutics, Icahn School of Medicine at Mount Sinai, New York, NY

Lizbeth Nunez (lizbeth.nunez@mssm.edu)
The Nash Family Center for Advanced Circuit Therapeutics, Icahn School of Medicine at Mount Sinai, New York, NY

Xiaosi Gu (xiaosi.gu@mssm.edu)
The Center for Computational Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY

Ignacio Saez (ignacio.saez@mssm.edu)
Department of Neurosurgery and Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY
Nash Family Department of Neuroscience and the Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY
The Nash Family Center for Advanced Circuit Therapeutics, Icahn School of Medicine at Mount Sinai, New York, NY

Abstract:

Counterfactual information is integral for optimal value-based decision-making. Counterfactuals are mental representations of alternative, hypothetical outcomes, which allow individuals to evaluate chosen and *unchosen* decisions. Aberrant counterfactual thinking is associated with multiple psychiatric disorders, like depression. While the neural encoding of counterfactual outcomes is well-defined, the neural and behavioral correlates of how counterfactual feedback affects future decision-making is unknown. Using human intracranial electrophysiology, we show the influence of counter-factual feedback on choice behavior is mediated by beta oscillations in the anterior insula and amygdala. These results provide a potential oscillatory mechanism for how previous counterfactual reward outcomes influence future decisions.

Keywords: counterfactual thinking; value-based choice; human intracranial electrophysiology; beta oscillations

Introduction

In value-based decision-making, humans evaluate choices based on their outcomes. In real-world decisions, choices have two types of outcomes, *actual* and *counterfactual*. Counterfactual thinking, the mental representation of alternative, hypothetical outcomes of unchosen decisions, is necessary to evaluate choices, and dysfunctional counterfactual thinking is a symptom of many psychiatric disorders (Howlett & Paulus, 2013). Despite the importance of counterfactual processing, the mechanism by which counterfactual information shapes future decisions is unknown.

To define the cognitive mechanism that integrates value signals from counterfactual rewards and current choice, we modeled choice latencies in a risky decision-making task that reveals actual and counterfactual outcomes. Task behavior was obtained from human epilepsy patients during intracranial electrophysiology (iEEG) recordings with uniquely high spatiotemporal resolution (Parvizi & Kastner, 2018). Anatomical targets for iEEG often include regions important for in value-based decision-making and representation of counterfactual information, including the orbitofrontal cortex (OFC) (Camille, 2004; Saez, 2018), anterior insula (aINS) (Howlett & Paulus, 2013), anterior cingulate cortex (ACC) (Hayden, 2009), amygdala (Amy) and hippocampus (HPC) (Coricelli, 2005), and dorsomedial (dmPFC) and ventromedial (vmPFC) prefrontal cortices (Van Hoeck, 2015). By combing cognitive models of choice behavior and human iEEG data, we identified the behavioral and neural mechanisms that facilitate processes underlying the integration of counterfactual information and choice-related value signals to generate risky decisions.

Methods

Patients performed a decision-making task with two choices: a risky gamble (*P(win)*=0.5) or a safe reward (*P(win)=1*) (n=150 trials; Fig 1A). After safe choices, patients saw the unchosen counterfactual gamble outcome. For every trial, we computed a counterfactual prediction error as the value difference between the actual and counterfactual outcomes: CPE = $V_A - V_C$.

Choice latencies related to CPE signals were quantified as the reaction time (RT) in the following trial. Trials with RT values < 300ms were excluded from analyses. We used linear mixed effects models with subject-level random effects to predict logRT on trial *t* from the CPE value on trial *t-1*. RT models included $choice_{t-1}$ and gamble EV_{t-1} as control covariates with variance inflation factor scores < 1.5.

We fit drift diffusion models (DDMs) using hierarchical sequential sampling modeling, a Bayesian parameter estimation tool that uses Markov chain Monte Carlo methods to evaluate hierarchical DDMs (Wiecki, 2013). We constructed hierarchical, linear DDMs to model drift rate, *v*, or decision bias, *z*, as a function of CPE with random subject-level slopes and intercepts (Navarro & Fuss, 2009). To evaluate and compare model fits, we used posterior predictive checks using leave-one-out cross-validation to estimate expected log predictive density (elpd-loo).

Neurophysiological data was preprocessed by removing artifacts from electrical noise and epileptic activity to improve signal-to-noise ratios. We used complex Morlet wavelet convolution with log spaced frequencies to extract oscillatory power from time-frequency representations (Cohen, 2008). Power estimates were extracted from the CPE outcome reveal epoch (3s) and z-scored to mean pre-trial baselines for cross-electrode analyses. We used mixed effects models with electrode-level random effects to predict trial-averaged beta (13-30Hz) power from CPEs (included control covariates, VIF < 1.5).

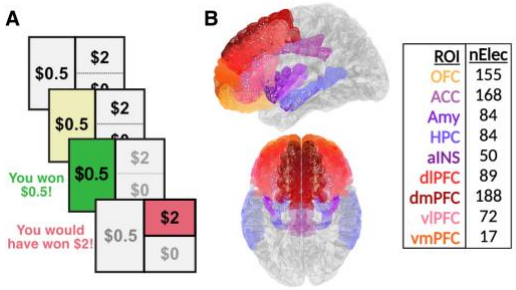

Figure 1 A. Example trial showing actual and counterfactual feedback following safe choice **B.** Localization of iEEG electrodes in ROIs (n=21 subjects)

Results

CPEs Decrease Choice Reaction Times To identify a putative mechanism for the effect of CPEs on choice computations, we modeled RTs as a function of CPE. We observed that RT_t was significantly modulated by CPE_{t-1} (β=0.019, z=1.97, p < 0.05). Individuals' response rates were dependent on the magnitude of preceeding CPEs; larger CPE_{t-1} values increased subsequent RT_t. We then tested whether the magnitude of patient-level CPE_{t-1} random effects was correlated with self-reported scores of depression symptom severity (Beck et al., 1996). Interestingly, the impact of CPEs on RTs was positively correlated with depression symptoms ($\rho(19) = 0.51$, $p < 0.02$).

CPEs Modulate Evidence Accumulation by Slowing Drift Rate Using hierarchical drift diffusion models, we investigated a potential mechanism for CPE modulation of decision processes. We hypothesized that the observed increases in choice latencies were the result of altered evidence accumulation during decision behavior following CPE encoding. We modeled the cognitive mechanisms underlying decisions as an evidence accumulation process driven by drift rate (*v*) and response bias (*z*). DDMs modeling CPEs as a function of drift rate, not bias, best represented subjects' choice latencies (elpd-loo_{diff} = 43.88). Large CPEs delayed subsequent choices by slowing evidence accumulation via decreasing drift rates.

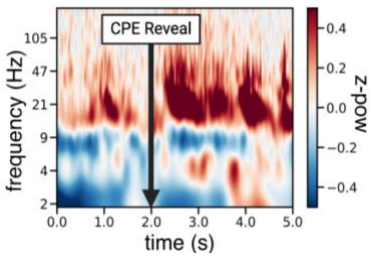

Figure 2. Power modulations following CPE reveal are specific to beta frequencies in the aINS (n=1subj; n=3elecs)

Beta Oscillations Mediate the Impact of CPEs on Decision-Making We hypothesized that the neural mechanisms underlying CPE modulation of subsequent decision processes are facilitated by beta oscillations. Oscillatory mechanisms involving beta frequencies underlie sensorimotor functions, top-down processing (Brincat & Miller, 2016), cognitive control mechanisms (Stoll, 2016), decision-making (Spitzer & Haegens, 2017), and, importantly, reward error encoding (Haufler, 2022). Preliminary visual examination of time-frequency representations revealed increased beta power activity aligned to CPE feedback (Fig 2). To ascertain the role of beta oscillatory power in CPE-mediated decision behaviors, we first validated the encoding of CPE signals in trial-by-trial beta power. We modeled electrode-level beta power as a function of CPEs and control covariates across all regions implicated in representing counterfactual information (Fig.1B). We found that electrode-level beta power modulations significantly driven by CPE (β=0.001, z= 2.62; p < 0.01). We evaluated this relationship in individual regions by aggregating the electrode-level random CPE slopes within ROIs. Surprisingly, ROIs most associated with counterfactual encoding (OFC, ACC) had minimal CPE encoding in beta power (Fig.3A). We hypothesized that beta power plays a unique role in counterfactual decision-making beyond encoding of CPE value.

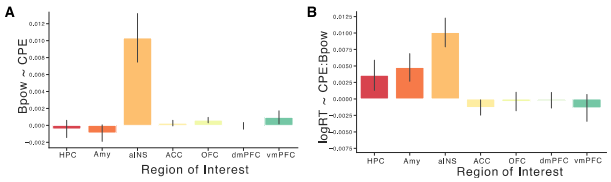

Figure 3A. CPEs modulate aINS beta (p<0.001) **B.** The CPE:Beta during feedback significantly predicts RT in the Amy (p<0.0005) and aINS (p<0.02).

Using electrode-level mixed effects models, we modeled reaction times RT_t as a function the CPE on the preceding trial, beta power, and the coefficient of interaction between CPE:beta power. We hypothesized that the strength of CPE:beta interactions during counterfactual processing would influence the degree to which CPEs modulates decision behavior. The CPE:Beta interaction from models fit on every electrode across ROIs did not significantly predict RTs (p<0.5; Fig.3B). However, post-hoc models fit separately for each ROI revealed the interaction between CPE:beta significantly predicts RTs following counterfactual feedback, specifically in the Amy and aINS. Therefore, the extent to which CPEs modulate drift rate is dependent on the degree of CPE feedback encoding in beta oscillations in the amygdala and anterior insula.

Discussion

Taken together, our results show that counterfactual feedback alters response to future decisions by slowing the rate of evidence accumulation. We identified the precise cognitive mechanism underlying choice latencies resulting from counterfactual feedback. Higher CPE value signals slow the rate of evidence accumulation, increasing observed reaction times. Importantly, this finding shows that counterfactual influences on RTs is the result of a complex cognitive process, rather than the downstream effect of psychomotor slowing. Finally, we defined a putative oscillatory mechanism in the amygdala and anterior insula that mediates the influence of CPEs on decisions though encoding of CPE values in beta oscillations during initial counterfactual feedback.

Acknowledgments

This work was supported by the National Science Foundation (graduate research fellowship to A Fink), and the National Institutes of Mental Health (grants number K01MH108815 and R01MH124763 to I.S.)

References

Beck, A. T., Steer, R. A., & Brown, G. (1996). *Beck Depression Inventory–II* [dataset].

Brincat, S. L., & Miller, E. K. (2016). Prefrontal Cortex Networks Shift from External to Internal Modes during Learning. *Journal of Neuroscience*, *36*(37), 9739–9754.

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.-R., & Sirigu, A. (2004). The Involvement of the Orbitofrontal Cortex in the Experience of Regret. *Science*, *304*(5674), 1167–1170.

Cohen, M. X. (2008). Assessing transient cross-frequency coupling in EEG data. *Journal of Neuroscience Methods*, *168*(2), 494–499.

Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A., & Dolan, R. J. (2005). Regret and its avoidance: A neuroimaging study of choice behavior. *Nature Neuroscience*, *8*(9), 1255–1262.

Haufler, D., Liran, O., Buchanan, R. J., & Pare, D. (2022). Human anterior insula signals salience and deviations from expectations via bursts of beta oscillations. *Journal of Neurophysiology*, *128*(1), 160–180.

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2009). Fictive Reward Signals in the Anterior Cingulate Cortex. *Science*, *324*(5929), 948–950.

Howlett, J. R., & Paulus, M. P. (2013). Decision-Making Dysfunctions of Counterfactuals in Depression: Who Might I have Been? *Frontiers in Psychiatry*, *4*.

Navarro, D. J., & Fuss, I. G. (2009). Fast and accurate calculations for first-passage times in Wiener diffusion models. *Journal of Mathematical Psychology*, *53*(4), 222–230.

Parvizi, J., & Kastner, S. (2018). Promises and limitations of human intracranial electroencephalography. *Nature Neuroscience*, *21*(4), 474–483.

Saez, I., Lin, J., Stolk, A., Chang, E., Parvizi, J., Schalk, G., Knight, R. T., & Hsu, M. (2018). Encoding of Multiple Reward-Related Computations in Transient and Sustained High-Frequency Activity in Human OFC. *Current Biology*, *28*(18), 2889-2899.e3.

Spitzer, B., & Haegens, S. (2017). Beyond the Status Quo: A Role for Beta Oscillations in Endogenous Content (Re)Activation. *Eneuro*, *4*(4), ENEURO.0170-17.2017.

Stoll, F. M., Wilson, C. R. E., Faraut, M. C. M., Vezoli, J., Knoblauch, K., & Procyk, E. (2016). The Effects of Cognitive Control and Time on Frontal Beta Oscillations. *Cerebral Cortex*, *26*(4), 1715–1732.

Van Hoeck, N. (2015). Cognitive neuroscience of human counterfactual reasoning. *Frontiers in Human Neuroscience*, *9*.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*.