Risk-sensitive reinforcement learning processes in the human brain drive impulsive choice

Rhiannon L. Cowan (rhiannon.cowan@utah.edu)
Department of Neurosurgery, University of Utah
Salt Lake City, UT 84132, USA

Tyler S. Davis (tyler.davis@hsc.utah.edu)
Department of Neurosurgery, University of Utah
Salt Lake City, UT 84132, USA

Bornali Kundu (bornali.kundu@gmail.com)
Department of Neurosurgery, University of Missouri
Columbia, MO 65212, USA

John D. Rolston (jrolston@bwh.harvard.edu)
Department of Neurosurgery, Brigham & Women's Hospital,
Boston, MA 02115, USA

Ben Shofty (ben.shofty@hsc.utah.edu)
Department of Neurosurgery, 175 N Medical Dr.,
Salt Lake City, UT 84132 USA

Shervin Rahimpour (shervin.rahimpour@hsc.utah.edu)
Department of Neurosurgery, University of Utah
Salt Lake City, UT 84132, USA

Elliot H. Smith (e.h.smith@utah.edu)
Department of Neurosurgery, University of Utah
Salt Lake City, UT 84132, USA

Abstract

We utilize standard and risk-sensitive reinforcement learning (RL) models to examine behavior and neural encoding of expected value and prediction error in 71 neurosurgical subjects during a risky decision-making task. We observed a behavioral trade-off between task performance and accuracy, specifically underpinned by negative prediction errors. Less impulsive choosers encoded RL model variables across brain regions to a greater extent, with greater encoding in the striatum, nucleus accumbens, and frontal cortices, compared to more impulsive choosers. More impulsive choosers' decision-making was dictated by negative prediction errors, leading to risk-aversive tendencies and, ultimately, suboptimal decision-making.

Keywords: reinforcement learning; risk sensitivity; impulsivity; decision-making

Introduction

Impulsive choice (IC) refers to the tendency to favor smaller, immediate, or more certain rewards over larger, delayed, or uncertain rewards (Hamilton et al., 2015) and is a prominent component of many psychiatric disorders (Huys et al., 2014). One approach by which individuals may make impulsive decisions is through risk aversion; by avoiding potential loss of reward and gaining instant gratification (Białaszek et al., 2015). In economics, risk is defined as the variance associated with an outcome (Niv et al., 2012), which may be examined via prediction error (PE) —a canonical signal of reinforcement learning (Preuschoff et al., 2006; Schultz, 2016). To avoid loss and gain instant reward, we posit that more impulsive individuals will exhibit risk-aversive tendencies, which will be observed via suboptimal task performance, related to differential learning rates associated with negative outcomes.

Methods

71 neurosurgical epilepsy patients underwent implantation of electrodes into the cortex and deep brain structures. Patients completed the Balloon Analog Risk Task (BART; Fig. 1A) while their brain activity was

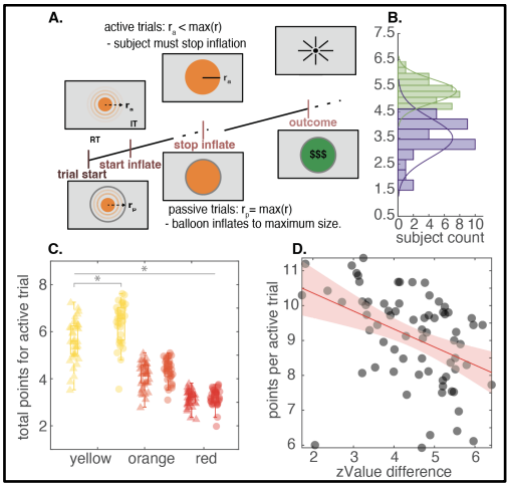

Figure 1: (A) BART schematic. r = reward (B) Histogram of IC score by group (MI = green, LI = purple) (C) Total points gained by balloon color (yellow, orange, red) between IC groups (MI = Δ, LI = o) (D) Regression between points gained and IC level.

recorded. During BART, subjects inflate and stop an artificial balloon to accumulate points. There are red, orange, and yellow balloons that have decreasing levels of risk, related to the balloon's potential inflation time (IT). Subject IC level was calculated as the difference between passive and active trial IT distributions for yellow balloons. A Gaussian mixture model classified subjects as more impulsive (MI, N = 37) or less impulsive (LI, N = 34) (Fig.1B). To examine neural correlates IC and reward, we examined broadband high frequency (HFA; 70-150Hz) activity from intracranial electrodes in drug-resistant epilepsy patients. Each subject's outcome-aligned HFA (correlated with population neuronal firing near the electrode (Manning et al., 2009); was modeled as a linear combination of temporal difference (TD) variables. BART allows us to measure trial-by-trial PE (Eq1), value expectation (VE) (Eq2), with optimal α's (Sutton & Barto, 2018) and a risk-sensitive model (Eq3) (Niv et al., 2012):

$$\textbf{Eq1. } \delta^t = r^t + V^t - V^{t-1}$$

$$\textbf{Eq2. } V^t = V^{t-1} + \alpha \cdot \delta(t_{outcome})$$

Eq3.

$$V^t = \begin{cases} V^{t-1} + \alpha^+ . \delta(t_{outcome}) \; if \; \delta(t_{outcome}) > 0, \\ V^{t-1} + \alpha^- . \delta(t_{outcome}) \; if \; \delta(t_{outcome}) < 0, \end{cases}$$

Results

Each subject averaged 233 ± 24 total trials and 83% ± 7 task accuracy. Balloon accuracy was higher for yellow (87%), and orange (88%) compared to red (61%) ($F(210)=139$, $p<.0001$), and points gained by

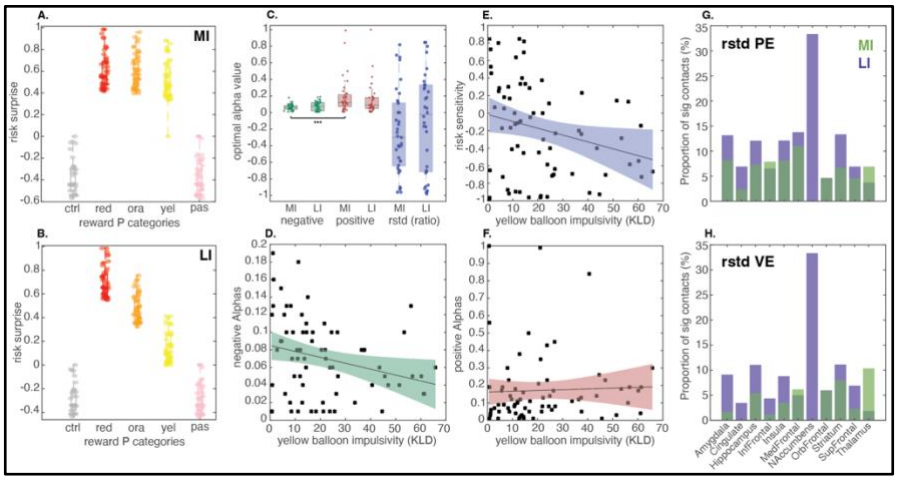

Figure 2: RSTD model. (A&B) Examples of MI/LI risk surprise by reward categories. (C) Optimal alphas for negPEs, posPEs, and Risk Sensitivity between IC groups. (D-F) Regressions between IC groups and negative, risk sensitivity, and positive α's. (G-H) RSTD PE & VE encoding between IC groups across ROIs.

balloon color were higher for yellow, orange, and red balloons, respectively ($F(210)=339$, $p<.0001$). Across all trials, MI choosers were more accurate than LI choosers ($Z=2.04$ $p=.041$), and rank-sum revealed that LI choosers were less accurate on yellow balloons compared to MI choosers ($Z=4.1$, $p<.0001$), but not other balloons. Across all trials, LI choosers gained more points compared to MI choosers ($Z=-3.6$, $p=.00036$), notably from statistically more yellow balloon points ($Z=-3.6$, $p=.00036$) (Fig.1C). Linking RSTD models outcomes to behavior, MI choosers showed differences between negative and positive α's calculated from the RSTD model ($Z=-3.3$, $p=0.0008$), but LI choosers didn't. Furthermore, impulsivity scores correlated to increased risk sensitivity scores ($t(71)=-2.2$, $p=.033$) and negative α's ($t(71)=-2.3$, $p=.025$). Across all electrode contacts, we observed the most encoding for the Reward PE (10.2%) and RSTD PE models (9.8%). For the RSTD model, we observed greater encoding of PE (19.3%) than VE (12.6%; $\chi2=42.8$, $p<.0001$). However, a group-level dichotomy revealed that LI choosers encoded more PE (8.3% vs 11.6%; $\chi2=15.9$, $p<.0001$) and more VE (5.3% vs 7.0%; $\chi2=6.4$, $p=.01$). For the overall PE TD model, we observed greater encoding of riskPE (20.0%) than rewardPE (14.4%; $\chi2=28.6$, $p<.0001$). Compared to MI choosers, LI choosers encoded more riskPE (5.8%% vs. 8.3%; $\chi2=12.5$, $p<.0001$) but similar rewardPE (9.0% vs. 11.5%). For the overall VE TD model, we observed greater encoding of rewardVE (15.4%) than riskVE (11.9%; $\chi2=13.1$, $p<.0001$). Neither MI nor LI groups encoded significantly more riskVE (8.1% vs. 6.9%) or rewardVE (6.4% vs. 5.8%), respectively. Compared to MI choosers, LI choosers encoded RSTD PE to a greater extent across most regions, notably the nucleus accumbens (0% vs. 33.3%), striatum (6.7% vs. 13.3%), insula (8.1% vs. 12.1%), and cingulate (2.4% vs. 6.8%). Similarly, LI choosers encoded RSTD VE to a greater extent across most regions, notably, the amygdala (1.6% vs. 9.1%), nucleus accumbens (0% vs. 33.3%), and striatum (8.0% vs. 11.1%), but not the thalamus (10.3% vs. 1.85%).

Discussion

In this study, we utilize an enormous intracranial dataset of 5167 electrodes to examine risk sensitivity and the neural underpinnings of IC using TD learning models with optimal α's. Our observation that MI choosers were more accurate but overall gained fewer points aligns with the impulsive tendency to opt for smaller, more immediate, and more certain rewards: stopping the balloon inflation early without reaping the full reward potential of the balloons. This tradeoff was particularly evident in the yellow balloon trials, which have the biggest reward potential, as MI choosers inflated balloons by 23.64% (~1.5s) less than LI choosers (Fig.1C). LI choosers tended to take more risks, which led them to perform more optimally on the task, while MI chooser's accuracy-point tradeoff suggests that they adopted a risk-aversive strategy. This behavioral finding is supported by differences in α's, showing that MI choosers may make decisions based on negative outcomes, which stifles potential reward gains, leading to suboptimal decision-making. Neurally, we observed that LI choosers encoded more PEs, which, in tandem with the differential behavioral strategies exhibited by the impulsivity groups, suggests risk sensitivity drives reward-seeking behaviors and may be modulated by impulsivity. These findings have implications for understanding the basis of decision-making, risk sensitivity, and impulsive choice.

Acknowledgements

This work was supporting by NIH funding: R01 (MH128187).

References

Białaszek, W., Gaik, M., McGoun, E., & Zielonka, P. (2015). Impulsive people have a compulsion for immediate gratification—Certain or uncertain. *Frontiers in Psychology*, *6*, 515. https://doi.org/10.3389/fpsyg.2015.00515

Hamilton, K. R., Mitchell, M. R., Wing, V. C., Balodis, I. M., Bickel, W. K., Fillmore, M., Lane, S. D., Lejuez, C. W., Littlefield, A. K., Luijten, M., Mathias, C. W., Mitchell, S. H., Napier, T. C., Reynolds, B., Schütz, C. G., Setlow, B., Sher, K. J., Swann, A. C., Tedford, S. E., … Moeller, F. G. (2015). Choice Impulsivity: Definitions, Measurement Issues, and Clinical Implications. *Personality Disorders*, *6*(2), 182–198. https://doi.org/10.1037/per0000099

Huys, Q. J. M., Tobler, P. N., Hasler, G., & Flagel, S. B. (2014). The role of learning-related dopamine signals in addiction vulnerability. In *Progress in Brain Research* (Vol. 211, pp. 31–77). Elsevier. https://doi.org/10.1016/B978-0-444-63425-2.00003-9

Manning, J. R., Jacobs, J., Fried, I., & Kahana, M. J. (2009). Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *Journal of Neuroscience*, *29*(43), 13613–13620.

Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural Prediction Errors Reveal a Risk-Sensitive Reinforcement-Learning Process in the Human Brain. *The Journal of Neuroscience*, *32*(2), 551–562. https://doi.org/10.1523/JNEUROSCI.5498-10.2012

Preuschoff, K., Bossaerts, P., & Quartz, S. R. (2006). Neural Differentiation of Expected Reward and Risk in Human Subcortical Structures. *Neuron*, *51*(3), Article 3. https://doi.org/10.1016/j.neuron.2006.06.024

Schultz, W. (2016). Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*, *18*(1), 23–32.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction.* MIT press.