Title: Distinct brain mechanisms linked to evidence accumulation and crime-type bias in juror decisions

Authors:

Jaime J. Castrellon[1,2], Shabnam Hakimi[2,3], Jacob M. Parelman[3,4], Lun Yin[2], Jonathan R. Law[1,2], Jesse A.G. Skene[1,2], David A. Ball[5], Artemis Malekpour[5], Donald H. Beskind[6], Neil Vidmar[6], John M. Pearson[1,2,7,8], J. H. Pate Skene[8,9]*, R. McKell Carter[3,10,11]*

*** -** Senior authors contributed equally to this work.

Affiliations:

1 - Department of Psychology and Neuroscience, Duke University

2 - Center for Cognitive Neuroscience, Duke University

3 - Institute of Cognitive Science, University of Colorado Boulder

4 - Annenberg School for Communication, University of Pennsylvania

5 - Malekpour & Ball Consulting (JuryWatch, Inc.)

6 - School of Law, Duke University

7 - Departments of Biostatistics & Bioinformatics, Duke University

8 - Department of Neurobiology, Duke University

9 - Initiative in Science and Society, Duke University

10 - Department of Psychology and Neuroscience, University of Colorado Boulder

11 - Electrical, Computer, & Energy Engineering, University of Colorado Boulder

Acknowledgements: The authors would like to thank Tal Yarkoni for his assistance in accessing the Neurosynth database. This work was supported by an Incubator Award from the Duke Institute for Brain Sciences, The Institute of Cognitive Science at CU Boulder to R.M.C, National Science Foundation (NSF) grant no. 1655445 to J.H.P.S, and by a career development award from the NIH Big Data to Knowledge Program (grant no. K01-ES-025442 to J.M.P). J.J.C was supported by an NSF Graduate Research Fellowship (grant no. NSF DGE-1644868). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Contributions: J.H.P.S., J.M.Pe, and R.M.C. conceived the project. All authors contributed to the task design. D.A.B., A.M., D.H.B., J.A.G.S, N.V., J.M.Pe, J.H.P.S., and R.M.C. wrote the crime scenarios and evidence modules. J.M.Pa, S.H., and R.M.C designed and coded the fMRI task, and J.H.P.S. ran the fMRI experiments. S.H., J.M.Pa., and J.J.C. processed the data. J.M.Pe. and L.Y. wrote/designed the computational models of behavior. J.J.C., J.M.Pe, J.H.P.S, and R.M.C. wrote/designed the computational models of fMRI data. J.J.C. analyzed the data. J.J.C., J.H.P.S, and R.M.C. wrote the paper. All authors approved the final version of the manuscript.

Data availability:

Unthresholded fMRI statistical maps can be viewed/downloaded from Neurovault: https://neurovault.org/collections/UADNVKNI/. Code and fMRI analysis pre-registration for hypothesis grouping can be viewed and downloaded from OSF: https://osf.io/rk92x/ and Github: https://github.com/jcastrel/juror_fmri

Conflicts of interest:

The authors have no conflicts of interest to report.

Keywords:

Decision making, law, evidence, social, fMRI

Abstract: Efforts to explain jury decisions have focused on competing models emphasizing utility, narrative, and social-affective mechanisms, but these are difficult to distinguish using behavior alone. Here, we use patterns of brain activation derived from large neuroimaging databases to look for signatures of the cognitive processes associated with models of juror decision making. We asked jury-eligible subjects to rate the strength of a series of criminal cases while recording the resulting patterns of brain activation. When subjects considered evidence, utility and narrative processes were both active, but cognitive processes associated with narrative models better explain the patterns of brain activation. In contrast, a biasing effect of crime type on perceived strength of the case was best explained by brain patterns associated with social cognition.

Main Text:

Complex social decisions encompass both evidence-based decision making and the influences of emotion, biases, and personal beliefs (*1*–*5*). Of these decisions, those made by jurors about the guilt or innocence of a criminal defendant are among the most consequential. As such, decision making in criminal trials provides a motivating and socially-relevant context for comparing the influences and brain mechanisms involved in evidence integration and bias in complex social decisions.

For juror decision making, two competing models from the general literature aim to explain how information is integrated to reach a decision: utility (*6*) and narrative (*7*). Utility is the foundational decision-theory model dealing with choices in the context of uncertainty. It is derived from a set of axioms and is simple and mathematically principled. In the context of criminal trials, utility theory assumes that jurors assess the cumulative weight of the evidence to estimate the probability that the defendant committed the crime, and compare that probability to a threshold based on expected harms or benefits (*8*–*11*). Narrative models, by contrast, propose that evidence integration relies on the extent to which the evidence can be assembled into a cohesive, compelling, and credible story (*1*, *12*). For jury decisions, narrative explanations are based on the extent to which the total body of information about a case can be structured into a coherent account and how closely the competing narratives offered by prosecution and defense match the juror's background knowledge, experience, and beliefs (*10*, *13*–*15*). Although most fields have relied heavily on the utility model, the narrative framework has been widely adopted for trial practice (*16*) while utility models play a role primarily in theoretical explanations.

In both the utility and narrative models, jurors may consider not only the evidence directly relevant to a decision but also factors not logically informative for deciding the facts of the case (*10*). We refer to those factors as extra-evidentiary. Extra-evidentiary factors can lead to biases based on racial or cultural stereotypes or desire to convict for serious crimes that evoke strong emotional or moral responses. Theories to explain extra-evidentiary influences (e.g., the severity of the crime) include: crime-type effects (generic prejudice (*10*, *17*) and leniency (*18*)); the liberation hypothesis (*10*, *19*); the role of emotions (*20*); moral judgment (*21*); and social assessment of the future impacts of the crime (*22*). These social-affective models treat extra-evidentiary information as a separate influence but an alternative is that the utility and narrative models of decision making explain both evidence and extra-evidentiary factors (*8*, *10*). Resolving these overlapping explanations requires a new approach.

Here, we use neural data to test for the involvement of cognitive processes associated with each model. We establish a priori hypotheses in which each model is framed in terms of a distinctive set of cognitive processes. Using whole brain decoding (*23*), we then compare patterns of brain activation associated with those cognitive processes to those observed during the processing of evidence and extra-evidentiary information in a criminal case. Finally, we extend the whole-brain decoding approach by comparing the goodness of fit for each model of juror decision making. This new approach allows us to resolve the contributions of each psychological model to a given decision.

To implement this approach, we collected functional magnetic resonance imaging (fMRI) data from jury-eligible participants as they viewed a series of simplified crime scenarios that vary in the type of crime and amount of evidence supporting guilt (*24*). Using this task (Fig. 1A), we previously showed that evidence and extra-evidentiary information about the crime independently influence participant ratings of the strength of the case against the accused (*24*). Consistent with the previous results, computational modeling of ratings by the current participants also distinguished the effects of different types of evidence and types of crime (Fig. 1B) on case strength. The extra-evidentiary influence of each type of crime on case strength is correlated with the severity of the crime, as measured by ratings of deserved punishment (Fig. 1C). We define the loading of each scenario along the dimension defined by this correlation (Fig. 1D) as crime-type bias. It captures the extent to which the description of a serious crime increases the rated strength of the case, independent of the evidence, akin to generic prejudice (*10*, *17*). To find the patterns of brain activation associated with these behavioral parameters, we modeled the neuroimaging data using separate parametric regressors of evidence accumulation and crime-type bias (see SI). We observed distinct constellations of activations associated with evidence accumulation (Fig. 2A, SI) and crime-type bias (Fig. 2B, SI).

As the next step in our analysis, we compared patterns of activation for evidence and crime-type bias to a large-scale database of published brain activations (*25*). Specifically, we compared thresholded brain-activation maps from the evidence and crime-type bias regressors to known activation maps from groups of cognitive processes associated with each psychological model of juror decision making. For each of these models, we defined a distinctive group of associated cognitive processes (Table S1). To limit user degrees of freedom utilized while defining model structure, we preregistered the list of topics associated with each model and submodel (https://osf.io/rk92x/). For the utility model, we included core cognitive topics associated with value, the consideration of risk, and mathematical calculation. For the narrative model, we included a series of submodels: 1. Experience, culture, and recall narrative (meant to encompass models associated with the consistency of a story and past experiences). This model was further subdivided into: 1a. culture and ideation bias and 1b. recall; 2. Working memory narrative (meant to capture the process of holding the pieces of a story together); and 3. Reading narrative (meant to capture models associated with processing fluency). For extra-evidentiary factors, we considered both affect and moral judgment. Within moral judgment, we included submodels for the process of applying morality and broader studies of social cognition. Hereafter, a model is described as significant if it is a better fit than 95% of models randomly drawn from three of the 177 topics not associated with one of the a priori models. We next describe how well each model is correlated with the patterns of brain activation associated with evidence accumulation and crime-type bias.

We sought to identify whether utility, narrative, social-affective, or some combination of, models explain the patterns of brain activation observed with increasing levels of evidence. To do so, we compared the patterns of brain activation for evidence accumulation to the patterns from the topics assigned to each model. Given the central role afforded to the narrative model in trial practice (*16*) and the foundational nature of the utility model in decision making more generally (*6*) we expected to find both narrative and utility models best explained the observed patterns of brain activation observed. Patterns of brain activation during evidence accumulation were correlated with utility and narrative topic maps but only very weakly with social-affective topic maps (Fig. 3A). The group of utility topic maps correlated with the pattern of brain activation during evidence accumulation (R^2=0.046, Fig. 3A). Of the contributing utility topics, probability calculation correlated best with evidence accumulation (R^2=0.038) and was the only submodel to explain a significant percentage of the variance (Fig. 3A dotted line, see SI for details). An aggregate model of narrative topics also correlated with patterns of brain activation during evidence accumulation (R^2=0.090, see Fig. 3A). Of the narrative submodels, both the recall narrative and reading narrative topic maps were significant (Fig. 3A, R^2=0.042 and R^2=0.065, respectively). An aggregate model including all social-affective topic maps (R^2=0.033) contained no submodels that explained a significant percentage of the variance. Thus, we found evidence for the role of specific aspects of the utility (probability calculation) and narrative (memory recall and reading) models in the consideration of evidence.

We next asked whether social-affective, utility, or narrative topic maps correlate with patterns of brain activation for a specific type of extra-evidentiary information, crime-type bias. To do so, we compared the patterns of brain activation for crime-type bias to the patterns from the topics assigned to each model. While utility and narrative models seek to explain both evidence and extra-evidentiary information, models specific to social-affective processing posit it as a separate process. If the utility and narrative models are correct, one or both may be associated with crime-type bias. If the social-affective models are correct, then those would explain crime-type bias but neither utility nor narrative would. We did, in fact, find that patterns of brain activation in response to crime-type bias (Fig 2B) were more strongly correlated with social-affective (R^2=0.050) and narrative (R^2=0.069) than utility topic maps (R^2=0.034, Fig. 3B). However, only topics from two specific submodels explain significant variance from the social-affective and narrative models, one in each. The social (theory of mind) component of the moral judgment submodel (R^2=0.045, Fig. 3B) drove effects in the social-affective model. In the narrative model, the fit was driven not by recall, as in the accumulation of evidence, but rather by the culture and ideation topic maps from the experience, culture, and recall submodel (R^2=0.054, Fig. 3B). No other submodel, explained significant variance (dotted line Fig. 3B, see SI). Thus, we found that crime-type bias was primarily associated with social and cultural topics but not emotion or recall.

Because the particular cognitive processes assigned to each model of juror decision making are subjective, we employed a data-driven approach to ask which of all 200 possible topics were most strongly associated with evidence accumulation and crime-type bias. To do so, we compared the patterns of activation from evidence accumulation and crime-type bias to reverse-inference maps for all 200 topic maps included in the Neurosynth database. We calculated the Pearson correlation between the thresholded parametric evidence accumulation map and the parametric crime-type-bias map and the 200 topic maps in Neurosynth (*25*) (SI methods). These findings confirmed, and further detailed, the role of narrative recall and mathematical calculation in evidence accumulation but also added a separate reasoning topic we had not included in our a priori models (Fig. 4A).

In a similar test, topics with the strongest correlations for crime-type bias were primarily centered on social cognition either directly (“mind, mental, social”, “social, interactions”, “type, joint”), indirectly (“gestures, abstract, race”, “body, bodies”), or anatomically (“junction, tpj, temporoparietal”, “temporal, sulcus, superior”) (Fig. 4B). Our a priori models assigned the two topics most correlated with crime-type bias (“gestures, abstract, race” and “mind, mental, social”) to experience, culture and recall and moral judgment respectively. Although this result includes two very different proposed models for decision making, it is similar to the a priori model results. Upon further examination, the two topic maps from this result heavily overlap in areas of the brain typically associated with social cognition, hinting at the prominent role social cognition plays in weighing extra-evidentiary information (Fig. S7).

Together, these findings suggest that distinct cognitive processes contribute to weighing evidence and extra-evidentiary effects on judgments of case strength, like crime-type bias. For evidence accumulation, we found engagement of both utility- and narrative-associated brain processes. The narrative model in aggregate explained almost twice the variance of the utility model. This supports more narrative influence while weighing evidence, although the larger number of topics in the narrative model may contribute to the observed difference. The process of weighing evidence showed little or no contribution from social-affective cognitive processes. We interpret both the cognitive processes associated with probabilistic reasoning (utility) (*26*) and recall of past experiences (narrative) (*27*) for evidence accumulation as supporting different aspects of the likelihood that a given story is true.

Crime-type bias looks quite different. First, social-affective processes play a greater role. Within the social-affective topics, social cognition, rather than affect (*28*) or traditional moral judgment (*29*), explains more of the variation. This contrasts with previous studies of the neural correlates of social punishment (*30–35*) and judgment bias (*36*). We have operationalized crime-type bias as the component of case strength ratings for each scenario that correlates with punishment (Fig. 1C-D). However, the neural correlates of crime-type bias effects on case strength are not the same as those previously linked to the punishment decision itself. Specifically, we found no indication that affective processes were associated with crime-type bias (though see (*37*)). In addition, although moral judgment has been associated with the TPJ, the topic maps associated with moral judgment include a number of other areas (e.g., basal ganglia) that were not present in the crime-type bias activations.

Second, for crime-type bias, the higher-level 'Narrative' model explains almost as much variance as it does above for evidence accumulation. However, the submodel that drives the contribution of 'Narrative' to crime-type bias is culture and ideation bias rather than recall and reading, as it was for evidence accumulation. The pattern of activation associated with culture and ideation bias is nearly identical to the social cognition topic from the social-affective model (Fig. S7). We note that social cognition is not necessarily a core component of narrative stories, it can be incorporated into broader models of narrative decision making (*1*). In fact, the contagious nature of a narrative has been tied to the role social value plays in making decisions about others (*38*). However, similar to past findings on difficult moral judgment (*39*), we observed brain activations primarily in the posterior aspects of the social-cognition network, not typically associated with social value. Instead, these posterior regions of the social cognition network are hypothesized to play a role in prediction (*40–42*). This role would be consistent with participants considering the consequences of future punishment and wrongful conviction.

In sum, we find neurally separable contributions of narrative, utility, and social-affective factors to juror decision making. For evidence accumulation, rather than supporting either a utility or narrative model alone, we interpret our results as supporting complementary aspects of the likelihood that the accusation is true. With regard to extra-evidentiary information, we find that social cognition can drive the biasing effect of crime-type, even in the absence of strong affect. In addition to informing best practices in the legal system, we offer this study as an example of a brain-decoding approach to elucidate complex decision making processes. A similar approach can be applied to quantifying the relative contributions of narrative, utility, and extra-evidentiary factors in a range of economic and social phenomena from asset pricing to the spread of fake news and market bubbles.

Fig. 1. A mock-juror task separates evidence accumulation and crime-type bias. During the task (**A**), participants read a criminal scenario paired with variable evidence from each of three types and rated the strength of the case and the recommended punishment severity. (**B**). Case-strength contributions from evidence independent of scenario and case-strength contributions from scenario independent of evidence. The mean ratings for each type of evidence and each scenario are shown as individual dots. Symbols represent mean effect size; error bars represent 95% credible intervals. Scenario effects are shown in rank order. Model estimates of the scenario on case strength and punishment independent of evidence were highly correlated (**C**). Crime-type bias is represented in the first principal component (PC) of a PCA on the model estimates (**D**).

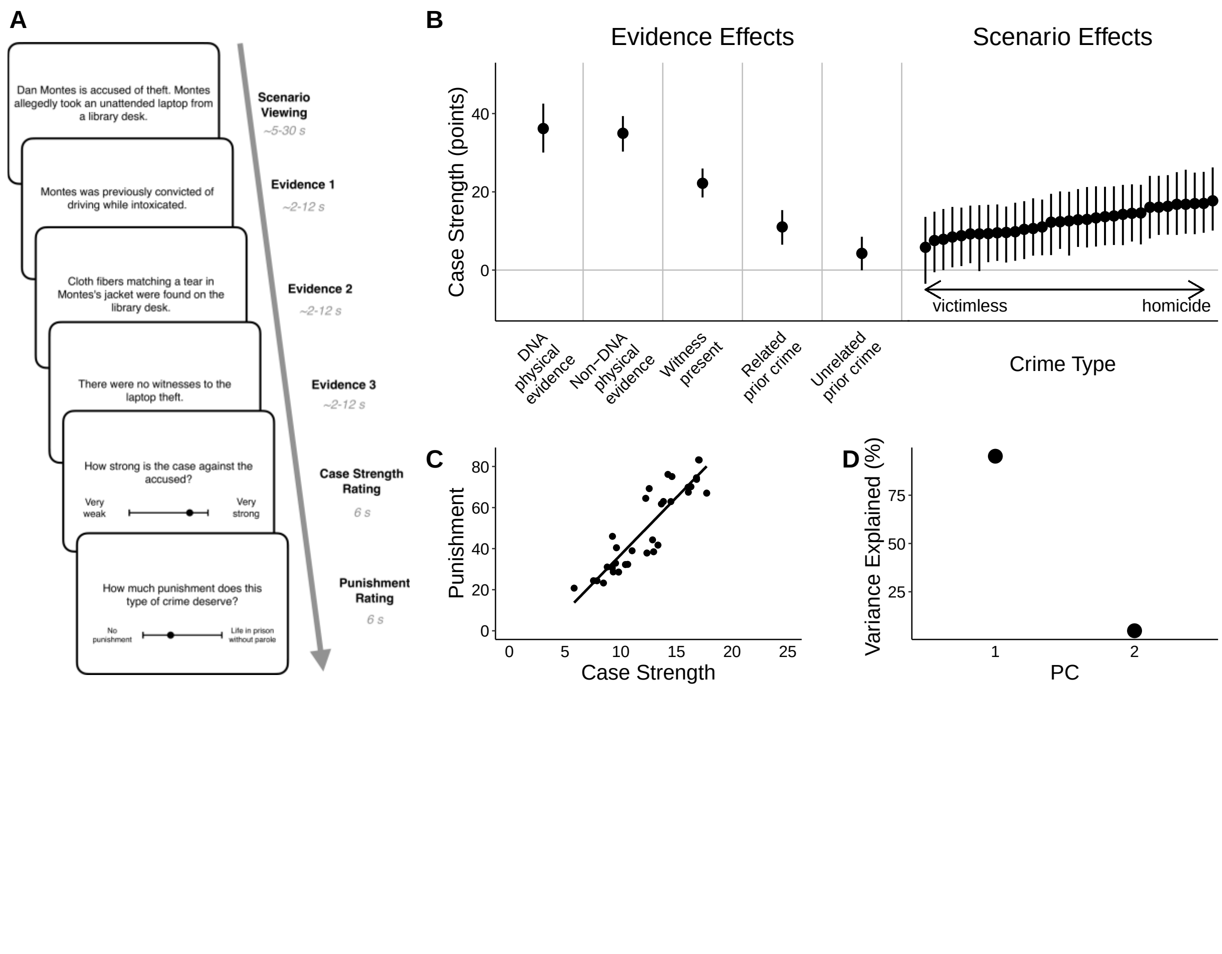


Fig. 2. Accumulation of evidence and crime-type bias produce different fMRI activation patterns. FMRI results are shown for (**A**) evidence accumulation during evidence presentation and (**B**) crime-type bias during scenario presentation. Results shown are corrected for multiple comparisons using a whole brain cluster-forming threshold Z > 2.3, cluster-corrected $p < 0.05$.

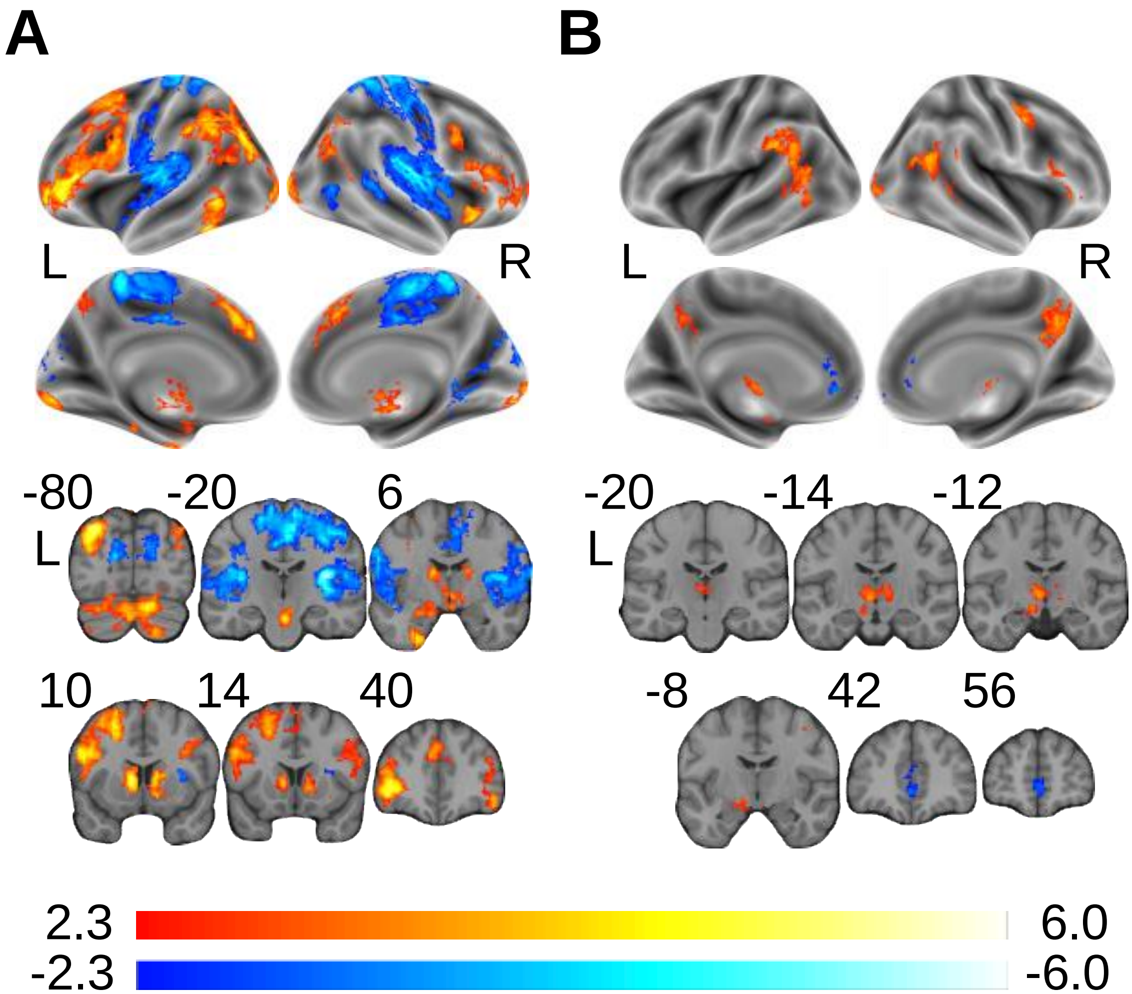


Fig. 3. Patterns of brain activation associated with evidence accumulation and crime-type bias support different decision-making models. Height of dendrograms shows the similarity (R^2) of thresholded parametric fMRI maps with Neurosynth topic models and submodels for (**A**) evidence accumulation and (**B**) crime-type bias. Red dotted lines are the 95 percentile of random models (see SI).

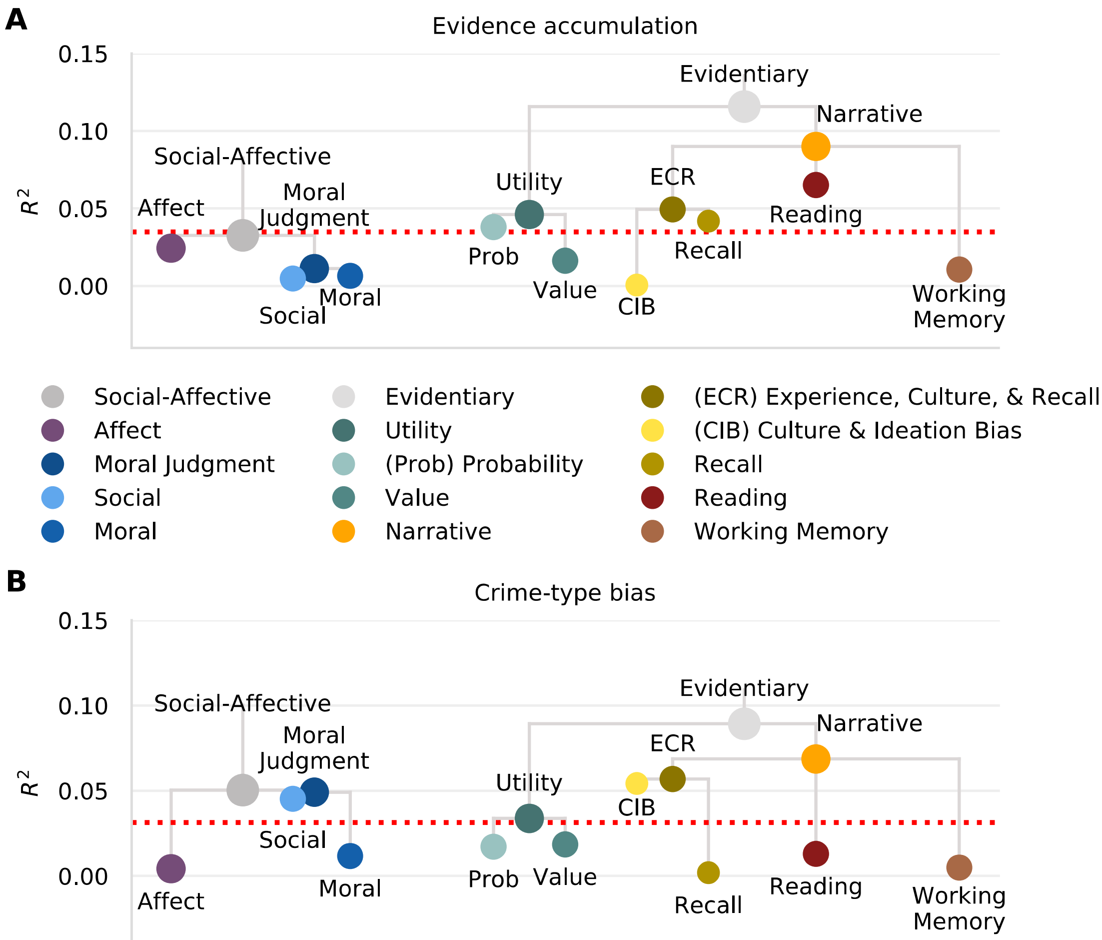
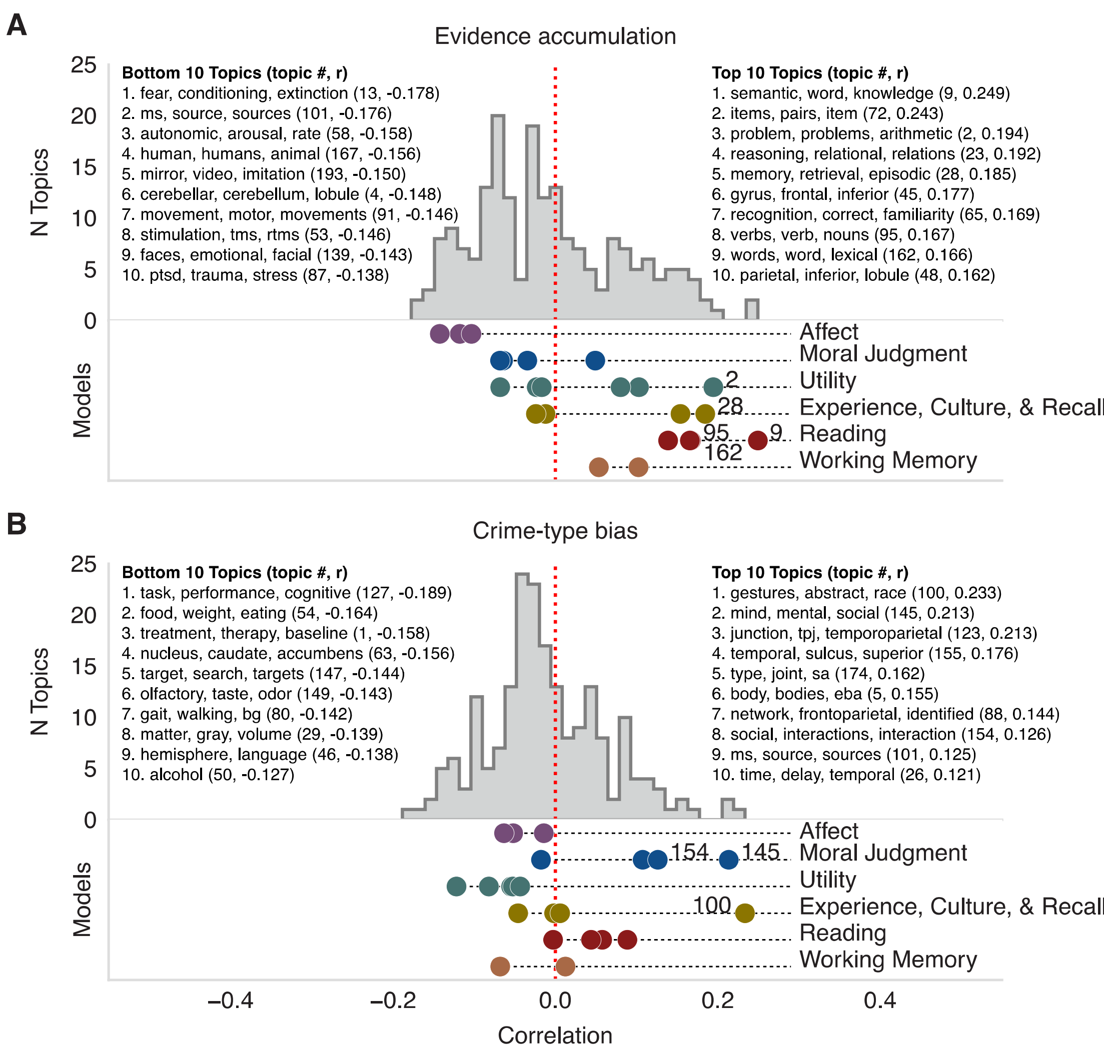


Fig. 4. Brain activation patterns associated with evidence accumulation and crime-type bias are associated with different cognitive processes. Histograms represent the distribution of spatial correlations between 200 Neurosynth topic maps and fMRI maps for (**A**) evidence accumulation and (**B**) crime-type bias. Below each histogram are points that correspond to the location of the correlation coefficients in the histogram for each topic included in each of the indicated decision-making models. The top and bottom 10 correlated topics from the entire set of 200 topics are indicated beside the histograms (top panels) and topics within each model that appear in the top 10 list are labeled beside their associated point for each model (bottom panels).



Bibliography

1. R. J. Shiller, Narrative Economics. *Am. Econ. Rev.* **107**, 967–1004 (2017).

2. R. B. Cialdini, N. J. Goldstein, Social influence: compliance and conformity. *Annu. Rev. Psychol.* **55**, 591–621 (2004).

3. S. Tremblay, K. M. Sharika, M. L. Platt, Social Decision-Making and the Brain: A Comparative Perspective. *Trends Cogn. Sci.* **21**, 265–276 (2017).

4. R. H. Thaler, C. R. Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Penguin, 2009).

5. A. Menon, Bringing cognition into strategic interactions: S trategic mental models and open questions. *Strategic Manage. J.* **39**, 168–192 (2018).

6. J. von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton University Press, 1944).

7. K. J. Arrow, Utilities, Attitudes, Choices: A Review Note. *Econometrica*. **26**, 1–23 (1958).

8. T. Connolly, Decision theory, reasonable doubt, and the utility of erroneous acquittals. *Law Hum. Behav.* **11**, 101–112 (1987).

9. H. R. Arkes, B. Shoots-Reinhard, R. S. Mayes, Disjunction Between Probability and Verdict in Juror Decision Making. *J. Behav. Decis. Mak.* **25**, 276–294 (2012).

10. D. J. Devine, *Jury Decision Making: The State of the Science* (NYU Press, 2012).

11. J. Kaplan, Decision Theory and the Factfinding Process. *Stanford Law Rev.* **20**, 1065–1092 (1968).

12. N. Pennington, R. Hastie, Evidence evaluation in complex decision making. *J. Pers. Soc. Psychol.* **51**, 242–258 (1986).

13. R. Hastie, The case for relative plausibility theory: Promising, but insufficient. *The International Journal of Evidence & Proof*. **23**, 134–140 (2019).

14. R. J. Allen, M. S. Pardo, Relative plausibility and its critics. *The International Journal of Evidence & Proof*. **23**, 5–59 (2019).

15. H. R. Arkes, J. P. Garske, *Psychological Theories of Motivation* (Brooks/Cole, 1982).

16. S. Lubet, J. C. Lore, *Modern Trial Advocacy Analysis & Practice: Fifth Edition (NITA)* (Wolters Kluwer, 5 edition., 2015).

17. N. Vidmar, Case studies of pre- and midtrial prejudice in criminal and civil litigation. *Law Hum. Behav.* **26**, 73–105 (2002).

18. N. L. Kerr, Severity of prescribed penalty and mock jurors' verdicts. *J. Pers. Soc. Psychol.* **36**, 1431–1442 (1978).

19. H. Kalven, H. Zeisel, T. Callahan, P. Ennis, *The american jury* (Little, Brown Boston, 1966).

20. C. Holloway, R. L. Wiener, The Role of Emotion and Motivation in Jury Decision-Making. *Criminal Juries in the 21st Century: Psychological Science and the Law*, 247 (2018).

21. J. Greene, J. Haidt, How (and where) does moral judgment work? *Trends Cogn. Sci.* **6**, 517–523 (2002).

22. J. W. Buckholtz, R. Marois, The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. *Nat. Neurosci.* **15**, 655–661 (2012).

23. R. Li, D. V. Smith, J. A. Clithero, V. Venkatraman, R. M. Carter, S. A. Huettel, Reason's Enemy Is Not Emotion: Engagement of Cognitive Control Networks Explains Biases in Gain/Loss Framing. *J. Neurosci.* **37**, 3588–3598 (2017).

24. J. M. Pearson, J. R. Law, J. A. G. Skene, D. H. Beskind, N. Vidmar, D. A. Ball, A. Malekpour, R. M. Carter, J. H. P. Skene, Modelling the effects of crime type and evidence on judgments about guilt. *Nat Hum Behav*. **2**, 856–866 (2018).

25. T. Yarkoni, R. A. Poldrack, T. E. Nichols, D. C. Van Essen, T. D. Wager, Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods*. **8**, 665–670 (2011).

26. S. a. Huettel, C. Jill Stowe, E. M. Gordon, B. T. Warner, M. L. Platt, Neural signatures of economic preferences for risk and ambiguity. *Neuron*. **49**, 765–775 (2006).

27. C. Baldassano, U. Hasson, K. A. Norman, Representation of Real-World Event Schemas during Narrative Perception. *J. Neurosci.* **38**, 9689–9699 (2018).

28. P. A. Kragel, K. S. LaBar, Decoding the Nature of Emotion in the Brain. *Trends Cogn. Sci.* **20**, 444–455 (2016).

29. L. Young, F. Cushman, M. Hauser, R. Saxe, The neural basis of the interaction between theory of mind and moral judgment. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 8235–8240 (2007).

30. M. Stallen, F. Rossi, A. Heijne, A. Smidts, C. K. W. De Dreu, A. G. Sanfey, Neurobiological Mechanisms of Responding to Injustice. *J. Neurosci.* (2018),

doi:10.1523/JNEUROSCI.1242-17.2018.

31. J. W. Buckholtz, C. L. Asplund, P. E. Dux, D. H. Zald, J. C. Gore, O. D. Jones, R. Marois, The neural correlates of third-party punishment. *Neuron*. **60**, 930–940 (2008).

32. O. Zinchenko, Brain responses to social punishment: a meta-analysis. *Sci. Rep.* **9**, 12800 (2019).

33. D. J. De Quervain, U. Fischbacher, V. Treyer, M. Schellhammer, U. Schnyder, A. Buck, E. Fehr, D. J.-F. de Quervain, The neural basis of altruistic punishment. *Science*. **305**, 1254–1258 (2004).

34. M. J. Crockett, J. Z. Siegel, Z. Kurth-Nelson, P. Dayan, R. J. Dolan, Moral transgressions corrupt neural representations of value. *Nat. Neurosci.* **20**, 879–885 (2017).

35. M. R. Ginther, R. J. Bonnie, M. B. Hoffman, F. X. Shen, K. W. Simons, O. D. Jones, R. Marois, Parsing the Behavioral and Brain Mechanisms of Third-Party Punishment. *J. Neurosci.* **36**, 9420–9434 (2016).

36. B. De Martino, D. Kumaran, B. Seymour, R. J. Dolan, Frames, biases, and rational decision-making in the human brain. *Science*. **313**, 684–687 (2006).

37. O. FeldmanHall, T. Dalgleish, R. Thompson, D. Evans, S. Schweizer, D. Mobbs, Differential neural circuitry and self-interest in real vs hypothetical moral decisions. *Soc. Cogn. Affect. Neurosci.* **7**, 743–751 (2012).

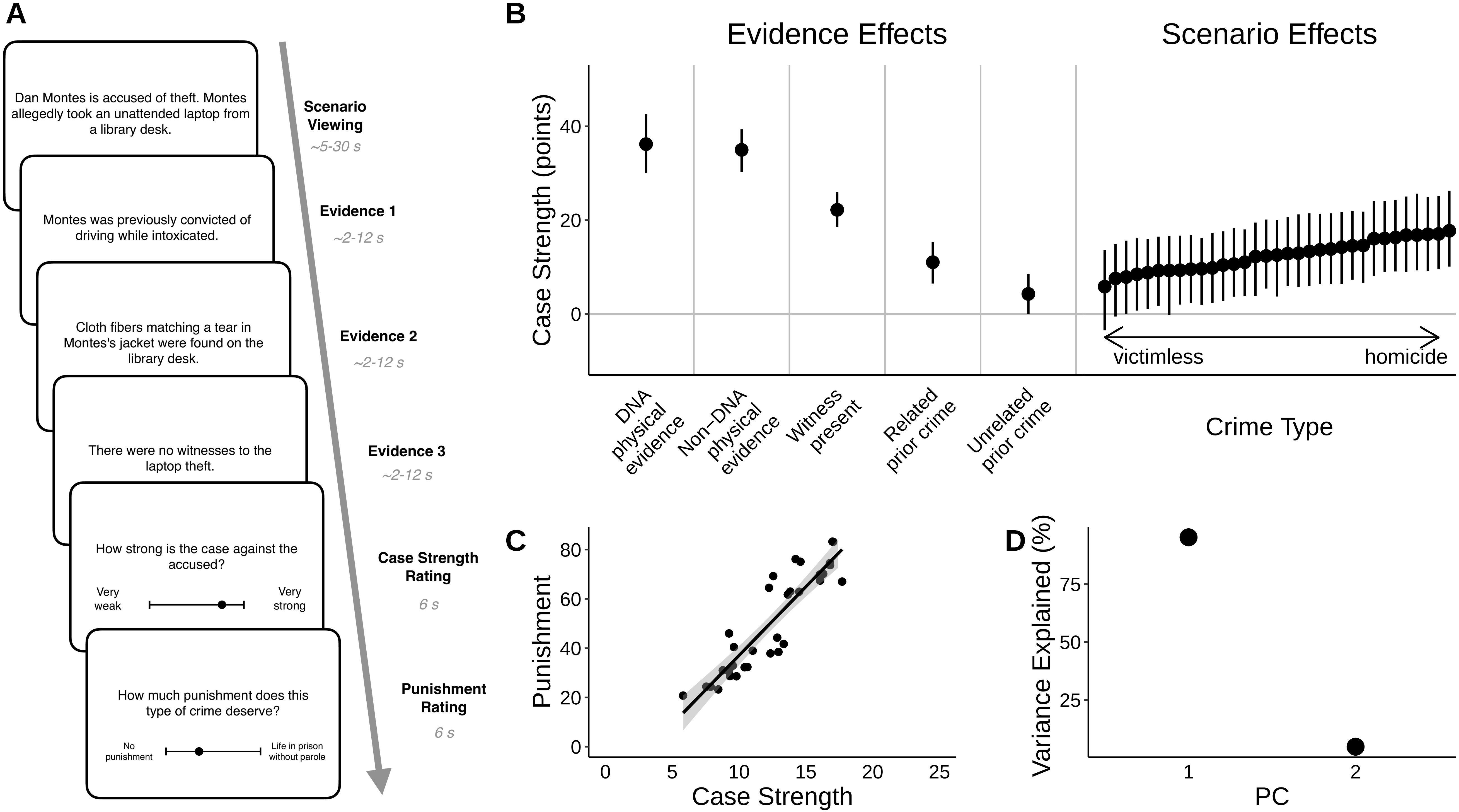
38. C. Scholz, E. C. Baek, M. B. O'Donnell, H. S. Kim, J. N. Cappella, E. B. Falk, A neural model of valuation and information virality. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2881–2886 (2017).

39. O. Feldmanhall, D. Mobbs, T. Dalgleish, Deconstructing the brain's moral network: dissociable functionality between the temporoparietal junction and ventro-medial prefrontal cortex. *Soc. Cogn. Affect. Neurosci.* **9**, 297–306 (2014).

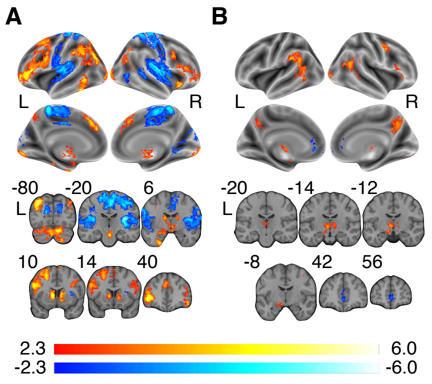
40. A. Soutschek, M. Moisa, C. C. Ruff, P. N. Tobler, The right temporoparietal junction enables delay of gratification by allowing decision makers to focus on future events. *PLoS Biol.* **18**, e3000800 (2020).

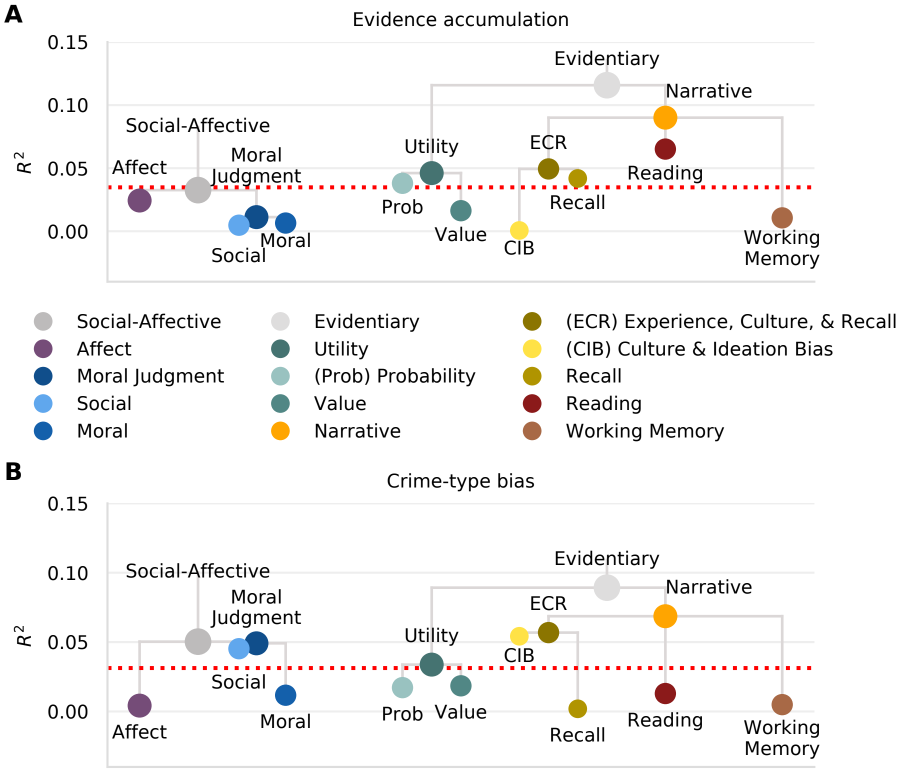
41. J. Koster-Hale, R. Saxe, Theory of mind: a neural prediction problem. *Neuron*. **79**, 836–848 (2013).

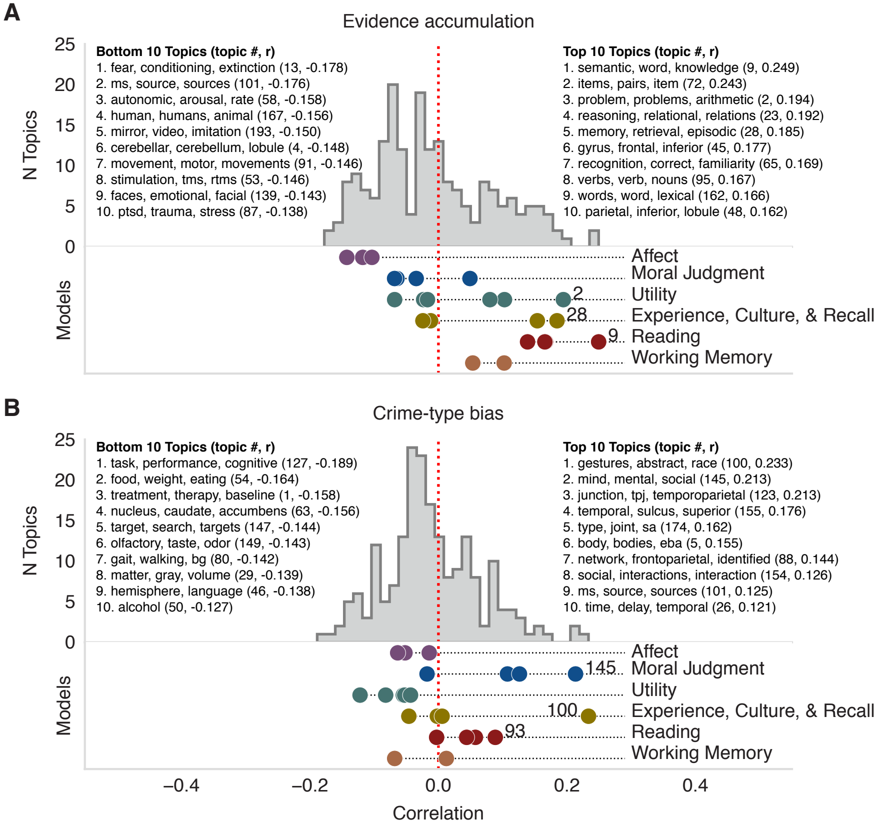
42. B. Park, D. Fareri, M. Delgado, L. Young, The role of right temporo-parietal junction in processing social prediction error across relationship contexts. *Soc. Cogn. Affect. Neurosci.* (2020), doi:10.1093/scan/nsaa072.



A
Dan Montes is accused of theft. Montes allegedly took an unattended laptop from a library desk.
Scenario Viewing
~5-30 s
Montes was previously convicted of driving while intoxicated.
Evidence 1
~2-12 s
Cloth fibers matching a tear in Montes's jacket were found on the library desk.
Evidence 2
~2-12 s
There were no witnesses to the laptop theft.
Evidence 3
~2-12 s
How strong is the case against the accused?
Very weak
Very strong
Case Strength Rating
6 s
How much punishment does this type of crime deserve?
No punishment
Life in prison without parole
Punishment Rating
6 s
B
Evidence Effects
Scenario Effects
Case Strength (points)
40
20
0
DNA physical evidence
Non−DNA physical evidence
Witness present
Related prior crime
Unrelated prior crime
victimless
homicide
Crime Type
C
Punishment
80
60
40
20
0
0
5
10
15
20
25
Case Strength
D
Variance Explained (%)
75
50
25
1
2
PC


A
B
L
R
L
R
-80
-20
6
L
10
14
40
-20
-14
-12
L
-8
42
56
2.3
6.0
-2.3
-6.0


A
Evidence accumulation
0.15
0.10
0.05
0.00
R^2
Social-Affective
Affect
Moral Judgment
Social
Moral
Utility
Prob
Value
Evidentiary
Narrative
ECR
CIB
Recall
Reading
Working Memory
Social-Affective
Affect
Moral Judgment
Social
Moral
Evidentiary
Utility
(Prob) Probability
Value
Narrative
(ECR) Experience, Culture, & Recall
(CIB) Culture & Ideation Bias
Recall
Reading
Working Memory
B
Crime-type bias
0.15
0.10
0.05
0.00
R^2
Social-Affective
Moral Judgment
Social
Affect
Moral
Utility
Prob
Value
Evidentiary
Narrative
ECR
CIB
Recall
Reading
Working Memory


A
Evidence accumulation
Bottom 10 Topics (topic #, r)
1. fear, conditioning, extinction (13, -0.178)
2. ms, source, sources (101, -0.176)
3. autonomic, arousal, rate (58, -0.158)
4. human, humans, animal (167, -0.156)
5. mirror, video, imitation (193, -0.150)
6. cerebellar, cerebellum, lobule (4, -0.148)
7. movement, motor, movements (91, -0.146)
8. stimulation, tms, rtms (53, -0.146)
9. faces, emotional, facial (139, -0.143)
10. ptsd, trauma, stress (87, -0.138)
Top 10 Topics (topic #, r)
1. semantic, word, knowledge (9, 0.249)
2. items, pairs, item (72, 0.243)
3. problem, problems, arithmetic (2, 0.194)
4. reasoning, relational, relations (23, 0.192)
5. memory, retrieval, episodic (28, 0.185)
6. gyrus, frontal, inferior (45, 0.177)
7. recognition, correct, familiarity (65, 0.169)
8. verbs, verb, nouns (95, 0.167)
9. words, word, lexical (162, 0.166)
10. parietal, inferior, lobule (48, 0.162)
N Topics
Models
Affect
Moral Judgment
Utility
Experience, Culture, & Recall
Reading
Working Memory
2
28
9
B
Crime-type bias
Bottom 10 Topics (topic #, r)
1. task, performance, cognitive (127, -0.189)
2. food, weight, eating (54, -0.164)
3. treatment, therapy, baseline (1, -0.158)
4. nucleus, caudate, accumbens (63, -0.156)
5. target, search, targets (147, -0.144)
6. olfactory, taste, odor (149, -0.143)
7. gait, walking, bg (80, -0.142)
8. matter, gray, volume (29, -0.139)
9. hemisphere, language (46, -0.138)
10. alcohol (50, -0.127)
Top 10 Topics (topic #, r)
1. gestures, abstract, race (100, 0.233)
2. mind, mental, social (145, 0.213)
3. junction, tpj, temporoparietal (123, 0.213)
4. temporal, sulcus, superior (155, 0.176)
5. type, joint, sa (174, 0.162)
6. body, bodies, eba (5, 0.155)
7. network, frontoparietal, identified (88, 0.144)
8. social, interactions, interaction (154, 0.126)
9. ms, source, sources (101, 0.125)
10. time, delay, temporal (26, 0.121)
145
100
93
−0.4
−0.2
0.0
0.2
0.4
Correlation