Sources of Philosophical Intuitions: Towards a Model of Intuition Generation— Bibliography

This document is a reference list accompanying the *Towards a Model of Intuition Generation* project. The list contains academic publications concerning the sampling mechanism introduced in the project. Synthesizing recent insights from cognitive psychology, the project assumes that intuitions are sampled from a heterogeneous and partitioned cognitive structure. Only a small number of intuitive claims can be sampled due to human cognitive limitations. Therefore, the model proposes that the sampling strategy people employ has a computationally efficient default marking the intuitive; that is, sampled claims are probable and valuable.

The list is divided into three major parts. The first part presents a range of empirical findings and theoretical implications reported in a growing body of literature suggesting a genuine role for value-guided judgments in non-normative, i.e., purely descriptive, questions. The second part is concerned with experimental and theoretical work on a shared sampling mechanism constituted by psychological representations of probability. The last part contains related findings such as framing effects and environmental factors that might influence intuition generation.

CONTENT

Value-Guided Judgment

In Philosophy

1. Adams, F., Steadman, A. (2004a). Intentional Action and Moral Considerations: Still Pragmatic. *Analysis*, *64*(3), 268–276. doi:10.1111/j.0003-2638.2004.00496.x
2. Adams, F., Steadman, A. (2004b). Intentional Action in Ordinary Language: Core Concept or Pragmatic Understanding? *Analysis*, *64*(282), 173–181. doi:10.1111/j.1467-8284.2004.00480.x
3. Alicke, M. D. (2000). Culpable Control and the Psychology of Blame. *Psychological Bulletin*, *126*(4), 556–574. doi:10. 1037/0033-2909.126.4.556
4. Alicke, M. D., Rose, D. (2010). Culpable Control or Moral Concepts? *Behavioral and Brain Sciences*, *33*(04), 330–331. doi:10.1017/S0140525X10001664
5. Alicke, M. D., Rose, D., Bloom, D. (2011). Causation, Norm Violation, and Culpable Control. *Journal of Philosophy*, *108*(12), 670–696. doi:10.5840/jphil20111081238
6. Alicke, M. D., Mandel, D. R., Hilton, D. J., Gerstenberg, T., Lagnado, D. A. (2015). Causal Conceptions in Social Explanation and Moral Evaluation: A Historical Tour. *Perspectives on Psychological Science*, *10*(6), 790–812. doi:10.1177/1745691615601888
7. Andow, J. (2017). Are Intuitions About Moral Relevance Susceptible to Framing Effects? *Review of Philosophy and Psychology*, *9*, 115–141. doi:10.1007/s13164-017-0352-5
8. Bartels, D. M., Medin, D. L. (2007). Are Morally Motivated Decision Makers Insensitive to the Consequences of Their Choices? *Psychological Science*, *18*(1), 24–28. doi:10.1111/j.1467-9280.2007.01843.x
9. Byrne, R. M. (2016) Counterfactual Thought: From Conditional Reasoning to Moral Judgment. *Annual Review of Psychology*, *67*, 135–157. doi:10.1146/annurev-psych-122414-033249
10. Cameron, C.D., Payne, B.K., Doris, J.M. (2013). Morality in High Definition: Emotion Differentiation Calibrates the Influence of Incidental Disgust on Moral Judgments. *Journal of Experimental Social Psychology*, *49*(4), 719–725. doi:10.1016/j.jesp.2013.02.014
11. Condon P., DeSteno, D. (2011). Compassion for One Reduces Punishment for Another. *Journal of Experimental Social Psychology*, *47*(3), 0–701. doi:10.1016/j.jesp.2010.11.016
12. Cova, F., Dupoux, E., Jacob, P. (2010). Moral Evaluation Shapes Linguistic Reports of Others' Psychological States, Not Theory-of-Mind Judgments. *Behavioral and Brain Sciences*, *33*(4), 334–335. doi:10.1017/s0140525x10001718
13. Cushman, F. (2008). Crime and Punishment: Distinguishing the Roles of Causal and Intentional Analyses in Moral Judgment. *Cognition*, *108*(2), 353–380. doi:10.1016/j.cognition.2008.03.006
14. Cushman, F., Young, L. (2011). Patterns of Moral Judgment Derive From Nonmoral Psychological Representations. *Cognitive Science*, *35*(6), 1052–1075. doi:10.1111/j.1551-6709.2010.01167.x
15. Egré, P., Cova, F. (2015). Moral Asymmetries and the Semantics of *Many*. *Semantics and Pragmatics*, *8*, 1–45. doi:10.3765/sp.8.13
16. Feltz, A., Cokely, E. T. (2008). The Fragmented Folk: More Evidence of Stable Individual Differences in Moral Judgments and Folk Intuitions. In B. C. Love, K. McRae, V. M. Sloutsky (Eds.), Proceedings of the 30th Annual Conference of the Cognitive Science Society (pp. 1771-1776). Austin, TX: Cognitive Science Society.
17. Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., Cohen, J. D. (2009). Pushing Moral Buttons: The Interaction Between Personal Force and Intention in Moral Judgment. *Cognition*, *111*(3), 364–371. doi:10.1016/j.cognition.2009.02.001
18. Guglielmo, S., Monroe, A. E., Malle, B. F. (2009). At the Heart of Morality Lies Folk Psychology. *Inquiry*, *52*(5), 449–466. doi:10.1080/00201740903302600
19. Haidt, J., Kesebir, S. (2010). Morality. In S. T. Fiske, D. T. Gilbert, G. Lindzey (Eds.), Handbook of Social Psychology, Volume 2 (5th Edition, pp. 797–832). Boston: McGraw-Hill.

20. Hitchcock, Ch., Knobe, J. (2009). Cause and Norm. *The Journal of Philosophy*, *106*(1), 587–612. doi:10.5840/jphil20091061128
21. Holton, R. (2010). Norms and the Knobe Effect. *Analysis*, *70*(3), 1–8. doi:10.1093/analys/anq037
22. Knobe, J. (2003a). Intentional Action and Side Effects in Ordinary Language. *Analysis*, *63*(3), 190–194. doi:10.1093/analys/63.3.190
23. Knobe, J. (2003b). Intentional Action in Folk Psychology: An Experimental Investigation. *Philosophical Psychology*, *16*(2), 309–325. doi:10.1080/09515080307771
24. Knobe, J. (2005). Theory of Mind and Moral Cognition: Exploring the Connections. *Trends in Cognitive Sciences*, *9*(8), 357–359. doi:10.1016/j.tics.2005.06.011
25. Knobe, J. (2006). The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology. *Philosophical Studies*, *130*, 203–231. doi:10.1007/s11098-004-4510-0
26. Knobe, J. (2010). Person as Scientist, Person as Moralist. *Behavioral and Brain Sciences*, *33*(4), 315–329. doi:10.1017/s0140525x10000907
27. Knobe, J., Burra, A. (2006). The Folk Concepts of Intention and Intentional Action: A Cross-Cultural Study. *Journal of Cognition and Culture*, *6*(1–2), 113–132. doi:10.1163/156853706776931222
28. Knobe, J., Fraser, B. (2008). Causal Judgment and Moral Judgment: Two Experiments. In W. Sinnott-Armstrong (Ed.), *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity* (Vol. 2). Cambridge, MA: The MIT Press.
29. Knobe, J., Szabo, Z. G. (2013) Modals With a Taste of the Deontic. *Semantics and Pragmatics*, *6*, 1–42. doi:10.3765/sp.6.1
30. Kominsky, J. F., Phillips, J. (2019). Immoral Professors and Malfunctioning Tools: Counterfactual Relevance Accounts Explain the Effect of Norm Violations on Causal Selection. *Cognitive Science*, *43*(11), e12792. doi:10.1111/cogs.12792
31. Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., Knobe, J. (2015). Causal Superseding. *Cognition*, *137*, 196–209. doi:10.1016/j.cognition.2015.01.013
32. Liu, B. S., Ditto, P. H. (2012). What Dilemma? Moral Evaluation Shapes Factual Belief. *Social Psychological and Personality Science*, *4*(3), 316-323. doi:10.1177/1948550612456045
33. Leslie, A. M., Knobe, J., Cohen, A. (2006). Acting Intentionally and the Side-Effect Effect: Theory of Mind and Moral Judgment. *Psychological Science*, *17*(5), 421–427. doi:10.1111/j.1467-9280.2006.01722.x
34. Levy, N. (2010). Scientists and the Folk Have the Same Concepts. *Behavioral and Brain Sciences*, *33*(4), 344. doi:10.1017/s0140525x10001809
35. Lombrozo, T., Uttich, K. (2010). Putting Normativity in Its Proper Place. *Behavioral and Brain Sciences*, *33*(4), 344–345. doi:10.1017/s0140525x10001810
36. Machery, E. (2008). The Folk Concept of Intentional Action: Philosophical and Experimental Issues. *Mind & Language*, *23*(2), 165–189. doi:10.1111/j.1468-0017.2007.00336.x
37. Mandelbaum, E., Ripley, D. (2010). Expectations and Morality: A Dilemma. *Behavioral and Brain Sciences*, *33*(4), 346–. doi:10.1017/s0140525x10001822
38. Mandelkern, M., Phillips, J. (2017) Force: Topicalization, Context-Sensitivity, and Morality. Retrieved from: https://osf.io/9shca/
39. Martin, K. (2009) *An Experimental Approach to the Normativity of "Natural."* [Paper presented at the Annual Meeting of the South Carolina Society for Philosophy, Rock Hill, South Carolina, February 27–28, 2009].
40. Michelin, C., Pellizzoni, S., Tallandini, M., Siegal, M. (2009). Evidence for the Side-Effect Effect in Young Children: Influence of Bilingualism and Task Presentation Format. *European Journal of Developmental Psychology*, *7*(6), 641–652. doi:10.1080/17405620902969989
41. Nadelhoffer, T. (2006). On Trying to Save the Simple View. *Mind & Language*, *21*(5), 565–586. doi:10.1111/j.1468-0017.2006.00292.x

42. Nichols, S. (2002). Norms With Feeling: Towards a Psychological Account of Moral Judgment. *Cognition*, *84*(2), 221–236. doi:10.1016/s0010-0277(02)00048-3

43. Nichols, S., Knobe, J. (2008). Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions. In J. Knobe, S. Nichols (Eds.), *Experimental Philosophy* (pp. 105–126). New York: Oxford University Press.

44. Pellizzoni, S., Siegal, M., Surian, L. (2009). Foreknowledge, Caring, and the Side-Effect Effect in Young Children. *Developmental Psychology*, *45*(1), 289–295. doi:10.1037/a0014165

45. Pettit, D., Knobe, J. (2009). The Pervasive Impact of Moral Judgment. *Mind & Language*, *24*(5), 586–604. doi:10.1111/j.1468-0017.2009.01375.x

46. Phillips, J., Cushman, F. (2017). Morality Constrains the Default Representation of What is Possible. *Proceedings of the National Academy of Sciences*, *114*(18), 4649–4654. doi:10.1073/pnas.1619717114

47. Phillips, J., Knobe, J. (2009). Moral Judgments and Intuitions About Freedom. *Psychological Inquiry*, *20*(1), 30–36. doi:10.1080/10478400902744279

48. Phillips, J., Luguri, J. B., Knobe, J. (2015). Unifying Morality's Influence on Non-Moral Judgments: The Relevance of Alternative Possibilities. *Cognition*, *145*, 30–42. doi:10.1016/j.cognition.2015.08.001

49. Pizarro, D. A., Helzer, E. G. (2010). Stubborn Moralism and Freedom of the Will. In R. F. Baumeister, A. R. Mele, K. D. Vohs (Eds.), *Free Will and Consciousness: How Might They Work?* (pp. 102–120). New York: Oxford University Press. doi:10.1093/acprof:oso/9780195389760.003.0007

50. Proft, M., Dieball, A., Rakoczy, H. (2019). What Is the Cognitive Basis of the Side-Effect Effect? An Experimental Test of Competing Theories. *Mind & Language*, *34*(3), 357–375. doi:10.1111/mila.12197

51. Schwitzgebel, E., Cushman, F. (2012). Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers. *Mind & Language*, *27*(2), 135–153. doi:10.1111/j.1468-0017.2012.01438.x

52. Semler, J., Henne, P. (2019). Recent Experimental Work on "Ought" Implies "Can". *Philosophy Compass*, *14*(9), e12619. doi:10.1111/phc3.12619

53. Shepherd, J. (2012). Action, Attitude, and the Knobe Effect: Another Asymmetry. *Review of Philosophy and Psychology*, *23*(2), 171–185. doi:10.1007/s13164-011-0079-7

54. Shtulman, A., Phillips, J. (2018). Differentiating "Could" From "Should": Developmental Changes in Modal Cognition. *Journal of Experimental Child Psychology*, *165*, 161–182. doi:10.1016/j.jecp.2017.05.012

55. Shtulman, A., Tong, L. (2013). Cognitive Parallels Between Moral Judgment and Modal Judgment. *Psychonomic Bulletin & Review*, *20*, 1327–1335 (2013). doi:10.3758/s13423-013-0429-9

56. Smetana, J. G., Rote, W. M., Jambon, M., Tasopoulos-Chan, M., Villalobos, M., Comer, J. (2012). Developmental Changes and Individual Differences in Young Children's Moral Judgments. *Child Development*, *83*(2), 683–696. doi:10.1111/j.1467-8624. 2011.01714.x

57. Sripada, C. S. (2010). The Deep Self Model and Asymmetries in Folk Judgments About Intentional Action. *Philosophical Studies*, *151*(2), 159–176. doi:10.1007/s11098-009-9423-5

58. Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., Ditto, P. H. (2009). The Motivated Use of Moral Principles. *Judgment and Decision Making*, *4*(6), 476–491.

59. Uttich, K., Lombrozo, T. (2010). Norms Inform Mental State Ascriptions: A Rational Explanation for the Side-Effect Effect. *Cognition*, *116*(1), 87–100. doi:10.1016/j.cognition.2010.04.003

60. Valdesolo, P., DeSteno, D. (2007). Moral Hypocrisy: Social Groups and the Flexibility of Virtue. *Psychological Science*, *18*(8), 689–690. doi:10.1111/j.1467-9280.2007.01961.x

61. Valdesolo, P., DeSteno, D. (2008). The Duality of Virtue: Deconstructing the Moral Hypocrite. *Journal of Experimental Social Psychology*, *44*(5), 0–1338. doi:10.1016/j.jesp.2008.03.010

62. Wysocki, T. (2020). Normality: A Two-Faced Concept. *Review of Philosophy and Psychology*, *11*(4), 689–716. doi:10.1007/s13164-020-00463-z

63. Young, L., Phillips J. (2011). The Paradox of Moral Focus. *Cognition*, *119*(2), 166–178. doi:10.1016/j.cognition.2011.01.004

In Cognitive Psychology

1. Armstrong, K., Schwartz, J. S., Fitzgerald, G., Putt, M., Ubel, P. A. (2002). Effect of Framing as Gain versus Loss on Understanding and Hypothetical Treatment Choices: Survival and Mortality Curves. *Medical Decision Making*, *22*(1), 76–83. doi:10.1177/0272989X0202200108
2. Bartlett, M. Y., DeSteno, D. (2006). Gratitude and Prosocial Behavior: Helping When It Costs You. *Psychological Science*, *17*(4), 319–325. doi:10.1111/j.1467-9280.2006.01705.x
3. Bocian, K., Myslinska-Szarek, K. (2021). Children's Sociomoral Judgements of Antisocial but not Prosocial Others Depend on Recipients' Past Moral Behaviour. *Social Development*, *30*(2), 396–409. doi:10.1111/sode.12480
4. Buon, M., Jacob, P., Loissel, E., Dupoux, E. (2013). A Non-Mentalistic Cause-Based Heuristic in Human Social Evaluations. *Cognition*, *126*(2), 149–155. doi:10.1016/j.cognition.2012.09.006
5. Chernyak, N., Kushnir, T. (2014). The Self as a Moral Agent: Preschoolers Behave Morally but Believe in the Freedom to Do Otherwise. *Journal of Cognition and Development*, *15*(3), 453–464. doi:10.1080/15248372.2013.777843
6. Decety, J., Wheatley, T. (2015). *The Moral Brain: A Multidisciplinary Perspective*. Cambridge: MIT Press.
7. DeSteno, D., Dasgupta, N., Bartlett, M. Y., Cajdric, A. (2004). Prejudice From Thin Air. The Effect of Emotion on Automatic Intergroup Attitudes. *Psychological Science*, *15*(5), 319–324. doi:10.1111/j.0956-7976.2004.00676.x
8. DeSteno, D., Petty, R. E., Rucker, D. D., Wegener, D. T., Braverman, J. (2004). Discrete Emotions and Persuasion: The Role of Emotion-Induced Expectancies. *Journal of Personality and Social Psychology*, *86*(1), 43–56. doi:10.1037/0022-3514.86.1.43
9. Druckman, J. N. (2001). Evaluating Framing Effects. *Journal of Economic Psychology*, *22*(1), 91–101. doi:10.1016/s0167-4870(00)00032-5
10. Falk, A., Szech, N. (2013). Morals and markets. *Science*, *340*(6133), 707–711. doi:10.1126/science.1231566
11. Freitas, J. D., DeScioli, P., Nemirow, J., Massenkoff, M., Pinker, S. (2017). Kill or Die: Moral Judgment Alters Linguistic Coding of Causality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/xlm0000369
12. Frisch, D. (1993). Reasons for Framing Effects. *Organizational Behavior and Human Decision Processes*, *54*(3), 399–429. doi:10.1006/obhd.1993.1017
13. Guglielmo, S. (2015). Moral Judgment as Information Processing: An Integrative Review. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.01637
14. Hamlin, J. K., Wynn, K., Bloom, P. (2007). Social Evaluation by Preverbal Infants. *Nature, 450*(7169), 557–559. doi:10.1038/nature06288
15. Iliev, R. I., Sachdeva, S., Medin, D. L. (2012). Moral Kinematics: The Role of Physical Factors in Moral Judgments. *Memory & Cognition*, *40*(8), 1387–1401. doi:10.3758/s13421-012-0217-1
16. Kalish, C. (2015). Normative Concepts. In E. Margolis, S. Laurence (Eds.), *The Conceptual Mind: New Directions in the Study of Concepts* (pp. 519–539). Cambridge, MA: MIT Press.
17. McElroy, T., Seta, J. J. (2003). Framing Effects: An Analytic–Holistic Perspective. *Journal of Experimental Social Psychology*, *39*(6), 0–617. doi:10.1016/s0022-1031(03)00036-2
18. Nolan, J., Schultz, P., Cialdini, R., Goldstein, N., Griskevicius, V. (2008). Normative Social Influence Is Underdetected. *Personality and Social Psychology Bulletin*, *34*(7), 913–923. doi:10.1177/0146167208316691
19. Rai, T. S., Fiske, A. P. (2011). Moral Psychology is Relationship Regulation: Moral Motives for Unity, Hierarchy, Equality, and Proportionality. *Psychological Review*, *118*(1), 57–75. doi:10.1037/a0021867
20. Reuter, K., Kirfel, L., Van Riel, R., Barlassina, L. (2014). The Good, the Bad, and the Timely: How Temporal Order and Moral Judgment Influence Causal Selection. *Frontiers in Psychology*, *5*, 1336. doi:10.3389/fpsyg.2014.01336

21. Stanovich, K. E., West, R. F. (2000). Individual Differences in Reasoning: Implications for the Rationality Debate? *Behavioral and Brain Sciences*, *23*(5), 645–665. doi:10.1017/s0140525x00003435

22. Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., Lerner, J. S. (2000). The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals. *Journal of Personality and Social Psychology*, *78*(5), 853–870. doi:10.1037/0022-3514.78.5.853

23. Tversky, A, Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, *211*(4481), 453–458. doi:10.1126/science.7455683

24. Valdesolo, P., DeSteno, D. (2006). Manipulations of Emotional Context Shape Moral Judgment. *Psychological Science*, *17*(6), 476–477. doi:10.1111/j.1467-9280.2006.01731.x

Probability-Guided Judgment

In Philosophy

1. Bear, A., Bensinger, S., Jara-Ettinger, J., Knobe, J., Cushman, F. (2020) What Comes to Mind? *Cognition 194*(104057). doi:10.1016/j.cognition.2019.104057
2. Bear, A., Knobe, J. (2017). Normality: Part Descriptive, Part Prescriptive. *Cognition*, *167*, 25–37. doi:10.1016/j.cognition.2016.10.024
3. Buckwalter, W. (2014). Gettier Made ESEE. *Philosophical Psychology*, *27*(3), 368–383. doi:10.1080/09515089.2012.730965
4. Egler, M., Ross, L. (2020). Philosophical Expertise Under the Microscope. *Synthese*, *197*, 1077–1098. doi:10.31234/osf.io/zxdbp
5. Egré, P. (2010). Qualitative Judgments, Quantitative Judgments, and Norm-Sensitivity. *Behavioral and Brain Sciences*, *33*(4), 335–336. doi:10.1017/s0140525x1000172x
6. Goodman, N. D., Tenenbaum, J. B., Gerstenberg, T. (2015). Concepts in a Probabilistic Language of Thought. In E. Margolis, S. Lawrence (Eds.), The Conceptual Mind: New Directions in the Study of Concepts (pp. 623–653). Cambridge, MA: MIT Press.
7. Gerstenberg, T., Goodman, N. D. (2012). Ping Pong in Church: Productive Use of Concepts in Human Probabilistic Inference. *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. Retrieved from: http://cicl.stanford.edu/papers/gerstenberg2012pingpong.pdf
8. Gerstenberg, T., Goodman, N. D., Lagnado, D., Tenenbaum, J. (2021). A Counterfactual Simulation Model of Causal Judgments for Physical Events. doi:10.31234/osf.io/7zj94
9. Gerstenberg, T., Icard, T. (2020). Expectations Affect Physical Causation Judgments. *Journal of Experimental Psychology: General*, *149*(3), 599–607. doi:10.1037/xge0000670
10. Halpern, J. Y., Hitchcock, C. (2015). Graded Causation and Defaults. *The British Journal for the Philosophy of Science*, *66*(2), 413–457. doi:10.1093/bjps/axt050
11. Henne, P., O'Neill, K., Bello, P., Khemlani, S., De Brigard, F. (2021). Norms Affect Prospective Causal Judgments. *Cognitive Science*, *44*. doi:10.1111/cogs.12931
12. Hitchcock C. R. (1996). The Role of Contrast in Causal and Explanatory Claims. *Synthese*, *107*(3), 395–419. doi:10.1007/bf00413843
13. Icard, T. F., Kominsky, J. F., Knobe, J. (2017). Normality and Actual Causal Strength. *Cognition, 161*, 80–93. doi:10.1016/j.cognition.2017.01.010
14. Lagnado, D. A., Gerstenberg, T., Zultan, R. (2013). Causal Responsibility and Counterfactuals. *Cognitive Science*, *37*(6), 1036–1073. doi:10.1111/cogs.12054
15. Livengood, J., Machery, E. (2007). The Folk Probably Don't Think What You Think They Think: Experiments on Causation by Absence. *Midwest Studies in Philosophy*, *31*(1), 107–127. doi:10.1111/j.1475-4975.2007.00150.x
16. Lombrozo, T. (2016). Explanatory Preferences Shape Learning and Inference. *Trends in Cognitive Sciences*, *20*(10), 748–759. doi:10.1016/j.tics.2016.08.001
17. Morris, A., Phillips, J., Gerstenberg, T., Cushman, F. (2019) Quantitative Causal Selection Patterns in Token Causation. *PLoS ONE*, *14*(8), e0219704. doi:10.1371/journal.pone.0219704
18. Morris, A., Phillips, J., Huang, K., Cushman, F. (2019). Habits of Thought Generate Candidate Actions for Choice. *PsyArXiv*. doi:10.31234/osf.io/2ayw3
19. Morris, A., Phillips, J., Icard, T., Knobe, J., Gerstenberg, T., Cushman, F. (2018). Judgments of Actual Causation Approximate the Effectiveness of Interventions. *PsyArXiv*. Retrieved from: https://psyarxiv.com/nq53z
20. Phillips, J., Morris, A., Cushman, F. (2019). How We Know What Not to Think. *Trends in Cognitive Sciences*, *23*(12), 1026–1040. doi:10.1016/j.tics.2019.09.007

21. Sytsma, J. (2019). Structure and Norms: Investigating the Pattern of Effects for Causal Attributions. Retrieved from: http://philsci-archive.pitt.edu/16626/
22. Sytsma, J. (2020). Causation, Responsibility, and Typicality. *Review of Philosophy and Psychology*. doi:10.1007/s13164-020-00498-2
23. Sytsma, J., Livengood, J., Rose, D. (2012). Two Types of Typicality: Rethinking the Role of Statistical Typicality in Ordinary Causal Attributions. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(4), 814–820. doi:10.1016/j.shpsc.2012.05.009
24. Turri, J. (2012). Is Knowledge Justified True Belief? *Synthese*, *184*(3), 247–259. doi:10.1007/s11229-010-9773-8

In Cognitive Psychology

1. Aragones, E., Gilboa, I., Postlewaite, A., Schmeidler, D. (2005). Fact-Free Learning. *American Economic Review*, *95*(5), 1355–1368. doi:10.1257/000282805775014308
2. Ashby, F.G., Maddox, W.T. (2005). Human Category Learning. *Annual Review of Psychology*, *56*(1), 149–178. doi:10.1146/annurev.psych.56.091103.070217
3. Bailenson, J. N., Shum, M. S., Atran, S., Medin, D. L., Coley, J. D. (2002). A Bird's Eye View: Biological Categorization and Reasoning Within and Across Cultures. *Cognition*, *84*(1), 1–53. doi:10.1016/s0010-0277(02)00011-2
4. Baker, C. L., Jara-Ettinger, J., Saxe, R., Tenenbaum, J. B. (2017). Rational Quantitative Attribution of Beliefs, Desires and Percepts in Human Mentalizing. *Nature Human Behaviour*, *1*(4), 0064. doi:10.1038/s41562-017-0064
5. Baker, C. L., Saxe, R., Tenenbaum, J. B. (2009). Action understanding as inverse planning. Cognition, 113(3), 329–349. doi:10.1016/j.cognition.2009.07.005
6. Barsalou, L. (1985). Ideals, Central Tendency, and Frequency of Instantiation as Determinants of Graded Structure in Categories. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *11*(4), 629–654. doi:10.1037/0278-7393.11.1-4.629
7. Bonawitz, E., Denison, S., Gopnik, A., Griffiths, T. L. (2014). Win-Stay, Lose-Sample: A Simple Sequential Algorithm for Approximating Bayesian Inference. *Cognitive Psychology*, *74*, 35–65. doi:10.1016/j.cogpsych.2014.06.003
8. Bramley, N. R., Gerstenberg, T., Tenenbaum, J. B., Gureckis, T. M. (2018). Intuitive Experimentation in the Physical World. *Cognitive Psychology, 105*, 9–38. doi:10.1016/j.cogpsych.2018.05.001
9. Burnett, R., Medin, D., Ross, N., Blok, S. (2005). Ideal is Typical. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *59*(1): 3–10. doi:10.1037/h0087453
10. Callaway, F., Griffiths, T. (2019). Attention in Value-Based Choice as Optimal Sequential Sampling. doi:10.31234/osf.io/57v6k
11. Cheng, P. W., Novick, L. R. (1990). A Probabilistic Contrast Model of Causal Induction. *Journal of Personality and Social Psychology*, *58*(4), 545–567. doi:10.1037/0022-3514.58.4.545
12. Cheng, P. W., Novick, L. R. (1992). Covariation in Natural Causal Induction. *Psychological Review*, *99*(2), 365–382. doi:10.1037/0033-295x.99.2.365
13. Costello, F., Watts, P. (2018). Invariants in Probabilistic Reasoning. *Cognitive Psychology*, *100*, 1–16. doi:10.1016/j.cogpsych.2017.11.003
14. Dasgupta, I., Gershman, S. J. (2021). Memory as a Computational Resource. *Trends in Cognitive Sciences*, *25*(3), 240–251. doi:10.1016/j.tics.2020.12.008
15. Dasgupta, I., Schulz, E., Gershman, S. J. (2017). Where Do Hypotheses Come From? *Cognitive Psychology*, *96*, 1–25. doi:10.1016/j.cogpsych.2017.05.001
16. Dasgupta, I., Schulz, E., Goodman, N.D., Gershman, S.J. (2018). Remembrance of Inferences Past: Amortization in Human Hypothesis Generation. *Cognition*, *178*, 67–81. doi:10.1016/j.cognition.2018.04.017
17. Dasgupta, I., Schulz, E., Tenenbaum, J. B., Gershman, S. J. (2020). A Theory of Learning to Infer. *Psychological Review*, *127*(3), 412–441. https://doi.org/10.1037/rev0000178
18. Denison, S., Bonawitz, E., Gopnik, A., Griffiths, T. L. (2013). Rational Variability in Children's Causal Inferences: The Sampling Hypothesis. *Cognition*, *126*(2), 285–300. doi:10.1016/j.cognition.2012.10.010
19. Dougherty, M. R., Hunter, J. E. (2003a). Hypothesis Generation, Probability Judgment, and Individual Differences in Working Memory Capacity. *Acta Psychologica*, *113*(3), 263–282. doi:10.1016/s0001-6918(03)00033-7
20. Dougherty, M. R., Hunter, J. (2003b). Probability Judgment and Subadditivity: The Role of Working Memory Capacity and Constraining Retrieval. *Memory & Cognition*, *31*(6), 968–982. doi:10.3758/bf03196449

21. Eidelman, C., Crandall, S. (2014). The Intuitive Traditionalist: How Biases for Existence and Longevity Promote the *Status Quo*. *Advances in Experimental Social Psychology*, *50*, 53–104. doi:10.1016/B978-0-12-800284-1.00002-3

22. Einhorn, H. J., Hogarth, R. M. (1986). Judging Probable Cause. *Psychological Bulletin*, *99*(1), 3–19. doi:10.1037/0033-2909.99.1.3

23. Epley, N., Gilovich, T. (2006). The Anchoring-and-Adjustment Heuristic. *Psychological Science*, *17*(4), 311–318. doi:10.1111/j.1467-9280.2006.01704.x

24. Evans, O., Stuhlmüller, A., Salvatier, J., Filan, D. (2017). Modeling Agents with Probabilistic Programs. Retrieved from: http://agentmodels.org

25. Fiedler, K. (2000). Beware of Samples! A Cognitive-Ecological Sampling Approach to Judgment Biases. *Psychological Review*, *107*(4), 659–676. doi:10.1037/0033-295x.107.4.659

26. Fincham, F. D., Jaspars, J. M. (1983). A Subjective Probability Approach to Responsibility Attribution. *British Journal of Social Psychology*, *22*(2), 145–161. doi:10.1111/j.2044-8309.1983.tb00575.x

27. Fleming, S. M., Daw, N. D. (2017). Self-Evaluation of Decision-Making: A General Bayesian Framework for Metacognitive Computation. *Psychological Review*, *124*(1), 91–114. doi:10.1037/rev0000045

28. Gershman, S. J., Vul, E., Tenenbaum, J. B. (2012). Multistability and Perceptual Inference. *Neural Computation*, *24*(1), 1–24. doi:10.1162/NECO_a_00226

29. Griffiths, T. L., Tenenbaum, J. B. (2006). Optimal Predictions in Everyday Cognition. *Psychological Science, 17*(9), 767–773. doi:10.1111/j.1467-9280.2006.01780.x

30. Griffiths, T.L., Vul, E., Sanborn, A.N. (2012). Bridging Levels of Analysis for Probabilistic Models of Cognition. *Current Directions in Psychological Science*, *21*(4), 263–268. doi:10.1177/0963721412447619

31. Halpern J.Y. (2011) Causality, Responsibility, and Blame: A Structural-Model Approach. In S. Benferhat, J. Grant (Eds), *Scalable Uncertainty Management. SUM 2011. Lecture Notes in Computer Science* (Vol 6929). Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-23963-2_1

32. Hawkins, G.E., Hayes, B.K., Heit, E. (2016). A Dynamic Model of Reasoning and Memory. *Journal of Experimental Psychology: General*, *145*(2), 155–180. doi:10.1037/xge0000113

33. Heit, E., Hayes, B. K. (2011). Predicting Reasoning from Memory. *Journal of Experimental Psychology: General*, *140*(1), 76–101. doi:10.1037/a0021488

34. Hilbert, M. (2012). Toward a Synthesis of Cognitive Biases: How Noisy Information Processing Can Bias Human Decision Making. *Psychological Bulletin*, *138*(2), 211–237. doi:10.1037/a0025940

35. Howe, R., Costello, F. (2020) Random Variation and Systematic Biases in Probability Estimation. *Cognitive Psychology*, 123. doi:10.1016/j.cogpsych.2020.101306

36. Icard, T. (2016). Subjective Probability as Sampling Propensity. *Review of Philosophy and Psychology*, *7*(4), 863–903. doi:10.1007/s13164-015-0283-y

37. Kahneman, D., Miller, D. T. (1986). Norm Theory: Comparing Reality to its Alternatives. *Psychological Review*, *93*(2), 136–153. doi:10.1037/0033-295X.93.2.136

38. Kirfel, L., Lagnado, D. (2018). Statistical Norm Effects in Causal Cognition. In T. T. Rogers, M. Rau, X. Zhu, C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 615–620). Madison, WI: Cognitive Science Society. Retrieved from: https://cogsci.mindmodeling.org/2018/papers/0132/0132.pdf

39. Kruschke, J.K. (2008). Models of Categorization. In R. Sun (Ed.), *The Cambridge Handbook of Computational Psychology* (pp. 267–301). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511816772.013

40. Lake, B., Ullman, T., Tenenbaum, J., Gershman, S. (2017). Building Machines That Learn and Think Like People. *Behavioral and Brain Sciences*, *40*, E253. doi:10.1017/S0140525X16001837

41. Lieder, F., Griffiths, T.L. (2020). Resource-Rational Analysis: Understanding Human Cognition as the Optimal Use of Limited Computational Resources. *Behavioral and Brain Sciences*, *43*, 1–60. doi:10.1017/S0140525X1900061X

42. Lieder, F., Griffiths, T.L., Hsu, M. (2018). Overrepresentation of Extreme Events in Decision Making Reflects Rational Use of Cognitive Resources. *Psychological Review*, 125(1), 1–32. doi:10.1037/rev0000074

43. Lieder, F., Griffiths, T. L., M. Huys, Q. J., Goodman, N. D. (2017a). The Anchoring Bias Reflects Rational Use of Cognitive Resources. *Psychonomic Bulletin & Review*, *25*(1), 322–349. doi:10.3758/s13423-017-1286-8

44. Lieder, F., Griffiths, T. L., Huys, Q. J. M., Goodman, N. D. (2017b). Empirical Evidence for Resource-Rational Anchoring and Adjustment. *Psychonomic Bulletin & Review*, *25*, 775–782. doi:10.3758/s13423-017-1288-6

45. Marchiori, D., Di Guida, S., Erev, I. (2015). Noisy Retrieval Models of Over- and Undersensitivity to Rare Events. *Decision*, *2*(2), 82–106. doi:10.1037/dec0000023

46. Moreno-Bote, R., Knill, D. C., Pouget, A. (2011). Bayesian Sampling in Visual Perception. *Proceedings of the National Academy of Sciences*, *108*, 12491–12496. doi:10.1073/pnas.1101430108

47. Oaksford, M., Chater, N. (2007). Bayesian Rationality: The Probabilistic Approach to Human Reasoning. Oxford: Oxford University Press.

48. Roberts, S., Gelman, S., Ho, A. (2016). So It Is, so It Shall Be: Group Regularities License Children's Prescriptive Judgments. *Cognitive Science*, *41*(Suppl 3), 576–600. doi:10.1111/cogs.12443

49. Roberts, S., Ho, A., Gelman, S. (2017). Group Presence, Category Labels, and Generic Statements Influence Children to Treat Descriptive Group Regularities as Prescriptive. *Journal of Experimental Child Psychology*, *158*, 19–31. doi:10.1016/j.jecp.2016.11.013

50. Sanborn, A.N., Chater, N. (2016). Bayesian Brains without Probabilities. *Trends in Cognitive Sciences*, *20*(12), 883–893. doi:10.1016/j.tics.2016.10.003

51. Sloman, S., Rottenstreich, Y., Wisniewski, E., Hadjichristidis, C., Fox, C.R. (2004). Typical Versus Atypical Unpacking and Superadditive Probability Judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 573–582. doi:10.1037/0278-7393.30.3.573

52. Sloman, S. A. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin*, *119*(1), 3–22. doi:10.1037/0033-2909.119.1.3

53. Smith, K. A., Hamrick, J. B., Sanborn, A. N., Battaglia, P. W., Gerstenberg, T., Ullman, T. D., Tenenbaum, J. B. (2021). *Probabilistic Models of Physical Reasoning*. Manuscript submitted for publication. Retrieved from: http://www.mit.edu/~k2smith/pdf/Smith_et_al-Probailistic_Physics.pdf

54. Stewart, N., Chater, N., Brown, G. D. (2006). Decision by Sampling. *Cognitive Psychology*, *53*(1), 1–26. doi:10.1016/j.cogpsych.2005.10.003

55. Stocker, A. A., Simoncelli, E. P. (2008). A Bayesian Model of Conditioned Perception. *Advances in Neural Information Processing Systems*. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4199208/

56. Suchow, J. W., Bourgin, D. D., Griffiths, T. L. (2017). Evolution in Mind: Evolutionary Dynamics, Cognitive Processes, and Bayesian Inference. *Trends in Cognitive Sciences*, *21*(7), 522–530. doi:10.1016/j.tics.2017.04.005

57. Tenenbaum, J. B., Kemp, C., Griffiths, T. L., Goodman, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, *331*(6022), 1279–1285. doi:10.1126/science.1192788

58. Thaker, P., Tenenbaum, J. B., Gershman, S. J. (2017). Online Learning of Symbolic Concepts. *Journal of Mathematical Psychology*, *77*, 10–20. doi:10.1016/j.jmp.2017.01.002

59. Thomas, R.P., Dougherty, M.R., Buttaccio, D.R. (2014). Memory Constraints on Hypothesis Generation and Decision Making. *Current Directions in Psychological Science*, *23*(4), 264–270. doi:10.1177/0963721414534853

60. Thomas, R.P., Dougherty, M., Sprenger, A.M., Harbison, J.I. (2008). Diagnostic Hypothesis Generation and Human Judgment. *Psychological Review*, *115*, 155–185. doi:10.1037/0033-295X.115.1.155

61. Trueblood, J. S., Busemeyer, J. R. (2011). A Quantum Probability Account of Order Effects in Inference. *Cognitive Science*, *35*(8), 1518–1552. doi:10.1111/j.1551-6709.2011.01197.x

62. Tversky, A., Kahneman, D. (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. doi:10.1126/science.185.4157.1124

63. Tversky, A., Kahneman, D. (1983). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review*, *90*(4), 293–315. doi:10.1037/0033-295x.90.4.293

64. Tversky, A., Koehler, D. J. (1994). Support Theory: A Nonextensional Representation of Subjective Probability. *Psychological Review*, *101*, 547–567. doi:10.1037/0033-295x.101.4.547

65. Tworek, C., Cimpian, A. (2016). Why Do People Tend to Infer "Ought" From "Is"? The Role of Biases in Explanation. *Psychological Science*, *27*(8): 1109–1122. doi:10.1177/0956797616650875

66. Ullman, T. D., Goodman, N. D., Tenenbaum, J. B. (2012). Theory Learning as Stochastic Search in the Language of Thought. *Cognitive Development*, *27*(4), 455–480. doi:10.1016/j.cogdev.2012.07.005

67. Voorspoels, W., Vanpaemel, W., Storms, G. (2011). A Formal Ideal-Based Account of Typicality. *Psychonomic Bulletin & Review*, *18*(5): 1006–1014. doi:10.3758/s13423-011-0122-9

68. Vul, E., Goodman, N., Griffiths, T.L., Tenenbaum, J.B. (2014). One and Done? Optimal Decisions From Very Few Samples. *Cognitive Science*, *38*(4), 599–637. doi:10.1111/cogs.12101

69. Vul, E., Pashler, H. (2008). Measuring the Crowd Within: Probabilistic Representations Within Individuals. *Psychological Science*, *19*(7), 645–647. doi:10.1111/j.1467-9280.2008.02136.x

70. Wang, Z., Solloway, T., Shiffrin, R. M., Busemeyer, J. R. (2014). Context Effects Produced by Question Orders Reveal Quantum Nature of Human Judgments. *Proceedings of the National Academy of Sciences*, *111*(26), 9431–9436. doi:10.1073/pnas.1407756111

Related Findings

In Philosophy

Demographic Differences

● ● ● Chandler, M. J., Proulx, T. (2008). Personal Persistence and Persistent Peoples: Continuities in the Lives of Individual and Whole Cultural Communities. In F. Sani (Ed.), Self-Continuity: Individual and Collective Perspectives (pp. 213–226). New York: Psychology Press.

Info:

Question: How can a person think she is still the same over time despite huge changes in beliefs, desires, values, aims, and so on? Essentialism: persistent persons are understood as possessing certain more or less abstract, but always defining bits or pieces that are said to defy time by remaining relentlessly the same. Narrativism: [presupposes] some unbroken chain of linked circumstances […] binds together life's various changing moments.

Results: 300 structured interviews. Among Aboriginal youth in Canada, 40% adopt the essentialist strategy, and 60% the narrativist strategy. Among non-Aboriginal youth: essentialism—80%, narrativism—20%.

Notable references:

• Chandler (2001). The time of our lives: Self-continuity in Native and non-Native youth. In Reese (Ed.), *Advances in child development and behavior*: Vol. 28 (pp.175-221). New York: Academic Press. • Chandler & Ball (1990). Continuity and commitment: A developmental analysis of the Identity formation process in suicidal and non-suicidal youth. In Bosma & Jackson (Eds.), Coping and self-concept in adolescence (pp. 149-166). New York: Springer-Verlag. • Chandler & Lalonde (1998). Cultural continuity as a hedge against suicide in Canada's First Nations. Transcultural Psychiatry, 35(2), 191-219. • Chandler & Lalonde (in press). Cultural continuity as a moderator of suicide risk among Canada's First Nations. In Kirmayer & Valaskakis (Eds.), The Mental Health of Canadian Aboriginal Peoples: Transformations, Identity, and Community. University of British Columbia Press. • Chandler, Lalonde, Sokol, & Hallett (2003). Personal persistence, identity development, and suicide: A study of Native and non-Native North American adolescents. Monographs of the Society for Research in Child Development, Serial No. 273, Vol. 68, No. 2.

● ● ● Feltz, A., Cokely, E. T. (2009). Do Judgments About Freedom and Responsibility Depend on Who You Are? Personality Differences in Intuitions About Compatibilism and Incompatibilism. *Consciousness and Cognition*, *18*(1), 342–350. doi:10.1016/j.concog.2008.08.001

Info:

Compatibilism = freedom and moral responsibility are compatible with determinism. Incompatibilism = ~compatibilism.

Extraversion. Communicative, sociable, energetic person who thrives on social contact and who does not regulate tightly his/her emotional reactions; extraverts enjoy social interaction, find social interacting rewarding, and actively seek out social interaction over being alone; extraverts are more skilled at decoding nonverbal communication than introverts; it is correlated with unique socially-minded judgment processes, scenario interpretations, and memory retrieval processes.

Method: A story about John, described in psychological terms (not neurological). Questions: 1. John's decision to kill his wife was **up to him**. 2. John decided to kill his wife **of his own free will** 3. John is morally **responsible** for killing his wife.

Results: on the scale of 1 (strongly agree)—7 (strongly disagree), the extraverts (bottom quartile) scored: up to him: 2, free will: 2.25, responsible: 1.5. The introverts (top quartile) scored: 3.75, 3.75, 3.

Explanation: Extraverts might be more likely to rely on judgment heuristics that are designed to facilitate social harmony (e.g. one's responsibility and freedom), regardless of whether everything about a person is said to be caused by previous events. That is, because extraverts are more socially-minded, they may dispropor-

tionately encode or perceive features of the individual in the scenario, perhaps drawing on examples from their own social experiences. In these ways, their intuitions may be proportionally less influenced by the deterministic features of the universe and more influenced by affective or social factors.

Individuals high in extraversion (e.g. socially-minded, outgoing, enthusiastic) are much more likely to judge that a person is free and responsible in a deterministic world than their non-extraverted counterparts. These results might partially explain why some philosophical debates are so intractable: People with different personalities, skills, and cognitive representations may simply have different intuitions about philosophical issues, perhaps as a product of different (conscious and unconscious) processes.

Questions: if they say that extraverts may disproportionately encode or perceive features of the individual in the scenario and be less influenced by the deterministic features of the universe, wouldn't that mean that extraversion causes a performance error and the introverts get the thing right? In short, maybe being an introvert is better, like being high in cognitive abilities (there are tests for high and low Need for Cognition but we don't want to say that NC is irrelevant—for a philosopher, it's better to be high in NC).

Notable references:

• Ericsson, A., & Simon, H. (1993). *Protocol analysis: Verbal reports as data* (Revised ed.). Cambridge: MIT Press. • Ashton, M., Lee, K., & Paunonen, S. (2002). What is the central feature of extraversion? Social attention versus reward sensitivity. *Journal of Personality and Social Psychology*, 83, 245–252. • Cokely, E. T., & Feltz, A. (submitted). Individual differences, judgments biases, and theory-of-mind: Deconstructing the intentional action side effect asymmetry. • Feltz, Cokely (2008). The fragmented folk: More evidence of stable individual differences in moral judgments and folk intuitions. In B. C. Love, K. & V. M. Sloutsky (Eds.), Proceedings of the 30th annual conference of the cognitive science society (pp. 1771–1776). Austin, TX: Cognitive Science Society. (those who displayed the largest judgment asymmetry were also high in the general personality trait extraversion) • Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37, 504–528. • Vargas, M. (2006). Philosophy and the folk: On some implications of experimental work for philosophical debates on free will. Journal of Cognition and Culture, 6, 239–254. • Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuition. *Nous*, 41, 663–685. (folk intuitions about moral responsibility can be influenced by (a) the "concreteness" or "abstractness" of the scenario and (b) by the emotional content of the scenario)

● ● ● Nadelhoffer, T., Kvaran, T., Nahmias, E. (2009). Temperament and Intuition: A Commentary on Feltz and Cokely. *Consciousness and Cognition*, *18*(1), 351–355. doi:10.1016/j.concog.2008.11.006

Info:

Feltz and Cokely's claim from the paper above: *the personality trait extraversion predicts people's intuitions about the relationship of determinism to free will and moral responsibility*. F&C didn't prove what they wanted because: (1) there were only 58 participants; (2) participants received only one vignette, which was (a) real-world (b) non-reductionist (c) deterministic (d) concrete, and (e) very high affect; (3) a truncated version of the Big-Five Personality Inventory was used that has a very limited two prompt extraversion sub-scale; and (4) only non-philosophers were used as participants.

The biggest problem is (2)—a single scenario. The more concrete and affect-laden a vignette is, the more likely participants will be to have compatibilist intuitions. Given that Feltz and Cokely explicitly state that the intuitions of extraverts may be proportionally less influenced by the deterministic features of the universe and more influenced by affective or social factors, they may have increased the likelihood that participants, regardless of their personality traits, would fail to appreciate the presence of determinism in the vignette. Moreover, they didn't use any control questions to check whether the participants understood the story.

Feltz and Cokely's results may not reveal anything about compatibilism specifically. Rather, extraversion could simply correlate with higher ratings of agreement in general to questions about control and responsibility, or to questions about (or questions that suggest) agents' deserving blame for their actions—the most likely indication of participants' "moral outrage" about an agent's immoral behavior. To test this interpretation, it would first be helpful to use a control scenario that did not involve determinism. The only conclusion we are

warranted in drawing at this point is that the moral valence of specific scenarios can sometimes influence the relationship between extraversion and compatibilist intuitions. Perhaps what F&C are tracking is not a direct relationship between extraversion and judgments of freedom and responsibility, but rather an indirect relationship between extraversion and a tendency to be susceptible to affective bias.

Notable references:

• John, Srivastava (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. John (Eds.), *Handbook of personality: Theory and research* (pp. 102–138). Guilford Press. • Nahmias, Coates, Kvaran (2007). Free will, moral responsibility, and mechanism: Experiments on folk intuitions. *Midwest Studies in Philosophy*, 31(1), 214–242.

● ● ● Schulz, E., Cokely, E. T., Feltz A. (2011). Persistent Bias in Expert Judgments About Free Will and Moral Responsibility: A Test of the Expertise Defense. *Consciousness and Cognition*, *20*(4), 0–1731. doi:10.1016/j.concog.2011.04.007

Info:

The expertise defense: if one has expert knowledge about some philosophical issue (e.g., the free will debate) the influences of extraneous features (e.g., personality) would not matter or would be dramatically reduced.

Intuitive judgments are prima facie better because the non-intuitive theories should provide an error theory for why folk thinks differently.

NEO-PI-R: each personality factor is itemized into 6 facets. Facets of extraversion: warmth, gregariousness, assertiveness, activity, excitement seeking, and positive emotion.

The experiment: 10 questions about the free-will debate to discriminate between experts and non-experts. Hypotheses: • extraversion will predict compatibilist judgments. • Experts will be more incompatibilists than non-experts. • Extraversion will continue to predict judgment biases in experts. Also checked for the Need for Cognition.

The story: Most respected neuroscientists […] specific chemical reactions and neural processes occurring in our brains. […] John decides to kill a shop owner because he needs money and does it. Once the specific thoughts, desires, and plans occur in John's mind, they will definitely cause his decision to kill a shop owner. Chemical reactions and neural processes in the story mixed with mind, thoughts, and desires in the actual case. The questions: 1 (absolutely agree)—7 (absolutely disagree): • John is morally responsible for his action. • John did it because of his own free will. • John's decision was up to him.

Results: • Warmth was the only facet related to compatibilistic judgments in a simple regression, 6% of the variance. • Greater philosophical knowledge was associated with more incompatibilist intuitions. The cognitive reflection task was not reliably related to intuitions in this scenario. • Among experts, warmth continued to predict a moderate amount of unique judgment variance (about 5% of variance).

Interpretation: Personality differences are heritable, and thus differences in philosophical views might be heritable as well. Three different ways for genotype and environment to covary. *Active*: one chooses environments according to one's genes. For example, each new generation of philosophers, because of their personality, finds some positions and examples intuitively appealing and adopts them. *Reactive*: where one's reaction to an environment is partially influenced by one's genes. For example, each new generation of philosophers will likely adopt a view that is consistent with previous philosophers with similar personalities because they react positively to views that are the same as their own. *Passive*: where one's environment is mostly formed by people that share in part the same genes. For example, if one is the child of a philosopher, one will probably not only share half of one's genes, but also some philosophy books.

Personality biases can lead to errors but also better accuracy. It is not clear why extraverts' intuitions should be trusted in Neo-Platonic projects, though it may be desirable for a jury to consist partially or completely of extraverts.

Notable references:

• Haidt, Koller, Dias (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65, 613–628. • Personality traits in philosophy: Cokely, Feltz, (2009). Individual differences, judgment biases, and theory-of-mind: Deconstructing the intentional action side effect asymmetry. *Journal of Research in Personality*, 43, 18–24. • Feltz., Cokely (2007). An anomaly in intentional action ascription: More evidence of folk diversity. In McNamara, Trafton, (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (p.1748). Mahwah, NJ: Erlbaum. • Feltz, Cokely (in press). The philosophical personality argument. • Cokely, Feltz (2009). Adaptive variation in folk judgment and philosophical intuition. *Consciousness and Cognition*, 18, 355–357. • Ethicists sometimes do not behave any better than non-ethicists: Schwitzgebel (2009). Do ethicists steal more books? *Philosophical Psychology*, 22, 711–725 • Schwitzgebel, Rust (2010). Do ethicists and political philosophers vote more often than other professors? *Review of Philosophy and Psychology*, 1, 189–199. • Schwitzgebel, Rust (2010). The moral behavior of ethicists: Peer opinion. *Mind*, 118, 1043–1069. • Sosa, E. (2007b). Intuitions: Their nature and epistemic efficacy. *Grazer Philosophische Studien*, 74, 51–67. • the extraversion scale from the NEO-PI-R: Costa, McCrae (1992). *NEO PI-R professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.

● ● ● Vaesen, K., Peterson, M., Van Bezooijen, B. (2013). The Reliability of Armchair Intuitions. *Metaphilosophy*, *44*(5), 559–578. doi:10.1111/meta.12060

Info:

The 1st study: 16 questions, English, Dutch and German philosophers asked. *Let suppose the following ___. Do you think that the thing S knows qualifies as knowledge?*

The biggest differences: *Karl knows that water is water*, native English: 75.4% (agree), Dutch: 49.6%. *Ludwig knows that on the 11th of September 2001 Osama bin Laden's eyes were either pink or some other color*, English: 75.1%, Dutch: 42.9%. *Hannah knows how to open her mouth*, English: 82.8%, Dutch: 42%. *Joseph knows that invisible objects aren't visible*, English 83.5%, Dutch: 66.1% (Dutch answers are always the most different).

Explanation of the results: in Dutch, German, and Swedish the words for *knowing* aren't related lexically to the word *knowledge*. Dutch *weten* refers to the psychological state of knowing (?), not implying any knowledge. Thus in the case of knowledge attribution, the epistemic standards for Germans and Scandinavians are higher than English speakers.

The 2nd study: more detailed stories (*Reinhardt is having a chemistry exam…*) to rule out the possibility that people fill in the details of an experiment differently. The average scores of Dutch philosophers are systematically significantly lower than English ones. Likert scale. Differences: *Patrick's finding his love of his life today*, E: 1.07, D: 1.33, *the way to the Museum of Modern art*, E: 2.68, D: 2.07, *the capital of Senegal*, E: 4.45, D: 4.

Notable references:

• Several excellent papers on "peer-disagreement" have emerged; for a concise overview see Christensen, D. (2009). Disagreement as Evidence: The Epistemology of Controversy. *Philosophy Compass* 4: 756-67.

● ● ● Feltz, A., & Cokely, E. T. (2008). The Fragmented Folk: More Evidence of Stable Individual Differences in Moral Judgments and Folk Intuitions. In B. C. Love, K. McRae, & V.M. Sloutsky (Eds.), Proceedings of the 30th Annual Conference of the Cognitive Science Society (pp. 1771-1776). Austin, TX: Cognitive Science Society.

Info:

Some philosophers take themselves to be analyzing philosophically interesting folk concepts. But there is not necessarily a "the folk," but instead that stable, identifiable groups of people express different intuitions and judgments about theoretically important topics.

People who are higher in the personality trait *openness to experience* tend to be more receptive to experience, less likely to reason in accordance with accepted societal standards, are more individualistic and do not take for

granted information passed on by authority. Individual differences would be related to those who also express non-objectivism about ethics. Those who are highly open to experience might be more likely to think that morals predominate in one's society are mistaken or otherwise flexible, and hence would be more open to the possibility that there is no single, correct ethical answer.

Notable references:

• the brief Big Five personality questionnaire Gosling, Rentfrow, Swann (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37, 504-528. • the brief self-control measure Tangney, Baumeister, Boone (2004). High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal of Personality*, 72(2), 271-322. • the operation-span (OSPAN) working memory task Turner, Engle (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, 28, 127-154. • the CRT Frederick (2005). Cognitive reflection and decision making. *Journal of Economic Perspective* 19(4), 25-41.

Order Effects

● ● ● Swain, S., Alexander, J., Weinberg, J. M. (2008). The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp. *Philosophy and Phenomenological Research*, *76*(1), 138–155. doi:10.1111/j.1933-1592.2007.00118.x

Info:

Epistemological reliabilism: a person's true belief that *p* counts as knowledge ≡ it is caused, or causally sustained, by a reliable cognitive process. The guy has no idea that the tempucomp has been inserted in his brain. Does he know the temperature? No, according to Lehrer. *According to standard practice, a philosophical claim is prima facie good to the extent that it accords with our intuitions, prima facie bad to the extent that it doesn't*. […] *even if one were to grant that, in principle, intuitions can be used as evidence in philosophy, these results suggest that, at this time, we cannot tell which intuitions can safely be deployed.*

Subjects who think that one knows in case of flipping a coin were filtered out. Likert scale: *Is it knowledge?* 1: strongly agree, 2: agree, 3: neutral, 4: disagree, 5: strongly disagree.

When the Truetemp after the Chemist (a clear case of knowledge), rated 2.4; when alone, 2.8, when after the Coinflip (a clear case of non-knowledge), rated 3.2.

Contextualism: the truth conditions vary according to the contexts in which they are uttered. The experiment's result: the context in which one applies the notion of knowledge must be empirically explored.

Notable references:

• David Lewis, "Scorekeeping in a Language Game", *Journal of Philosophical Logic* Vol. 8 (1979): pp. 339-59.

● ● ● Wright, J. C. (2010). On Intuitional Stability: The Clear, the Strong, and the Paradigmatic. *Cognition*, *115*(3), 491–503. doi:10.1016/j.cognition.2010.02.003

Info:

There are introspective methods for distinguishing between stable and unstable intuitions: confidence and belief strength.

Experiment 1. Cases used: *Truetemp*, *Coin-flip*, *Fake-Barn*, *Testimony* (a professor and a scientific journal). Additional question: how confident the respondent was (0-5). 55% say Truetemp is a case of knowledge if it follows Coin-flip, 40% if it follows Testimony, and 26% if it follows Fake-Barn. 40% say Fake-Barn is knowledge if it follows Testimony, 39% if Coin-Flip, 59% if True-temp. Testimony and Coin-flip aren't vulnerable to the order effect, True-temp and Fake-Barn are. Confidence in the stable cases was (mean) 4.4 for Coin-Flip and 4.5 for Testimony, in the unstable cases 3.9 for both True-temp and Fake-Barn. Thus, participants' low confidence ≡ instability.

In case of low confidence, the participants turn elsewhere, such as to the case that they had just previously considered, for information that would help to determine their judgment. Thus, participants who saw Coin-Flip first were more inclined to assess True-temp as knowledge because it looks a lot more like knowledge than a special feeling. For those who saw Testimony, it looks a lot less like knowledge than testimony from a scientific journal.

Experiment 2. Moral cases added: *Break-promise* (a guy doesn't meet with his girlfriend because his grandfather was taken to a hospital), *Sell-iPod* (Suzy wants to Laura's iPod which Laura thinks she lost), *Hide-Bombers* (During the war, hiding Jewish neighbors who bombed a schoolyard). New epistemic cases: *Perception* (one sees a red apple and thinks it's an apple) and *Farmer* (a farmer thinks his cow in the field. It is but actually he saw something else and took it for a cow). Additional questions: how strongly the participants believed, how confident whey were (ϱ>.85 between both questions), paradigmaticity—whether the peers agree on the issue.

Results: Break-promise was judged not wrong (97%), Sell-iPod was judged wrong (100%). Truetemp and Hide-Bombers were instable. 55% judged Hide-Bombers wrong if it followed Sell-iPod, 32% when it followed Break-Promise. The stable cases were perceived as paradigmatic, means: 5.6-6.5. The unstable cases were judged less paradigmatic: 4.9 for True-temp and 5.2 for Break-Promise. The farmer wasn't unstable (no order effect), but there was little consensus (1/3rd said he knew, 2/3rd said he didn't); its paradigmaticity 5.3 fell in between paradigmaticity levels of epistemic stable (5.6-6.5) and unstable (4.9) cases; the same with the belief strength.

Conclusions: there are non-experimental ways to predict intuitional instability: by confidence/strength of belief and by perceived consensus. Paradigmaticity increases stability because paradigmatic cases represent clear instances of concepts. Thus, people can introspectively track biases focusing on strength of their beliefs and consensus regarding the issue.

Because philosophers have also strong intuitions about non-paradigmatic cases, we should believe them—they are trained in thinking about concepts. Though, if the intuition is clear and strong it doesn't mean its content is true.

Notable references:

• Arkes, H. R. (2001). Overconfidence in judgmental forecasting. In Armstrong (Ed.), Principles of forecasting: A handbook for researchers and practitioners (pp. 495-516). Boston: Kluwer Academic. • Einhorn, Hogarth (1978). Confidence in judgment: Persistence of the illusion of validity. Psychological Review, 85, 395-416. • Epstein, Lipson, Holstein, Huh (1992). Irrational reaction to the negative outcome: Evidence for two conceptual systems. Journal of Personality and Social Psychology, 62, 328-339. • Gendler (2007). Philosophical thought experiments, intuitions, and cognitive equilibrium. Midwest Studies in Philosophy, 31:1, 68-89. • Griffin, Tversky, A. (1992). The weighing of evidence and the determinants of confidence. Cognitive Psychology, 243, 411-435. • Haidt, Joseph (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. Daedalus, 55-66. • Kahneman, Klein (2009). Conditions for intuitive expertise: A failure to disagree. American Psychologist, 64:6, 515-526. • Kahneman, Tversky (1982). On the study of statistical intuitions. Cognition, 11, 123-141. • King, Appleton, 1997. Intuition: A critical review of the research and rhetoric. Journal of Advanced Nursing, 26, 194-202 • Krosnick, Petty (1995). Attitude strength: An overview. In Petty, Krosnick (Eds.), Attitude strength: Antecedents and consequences (pp. 1-24). Hillsdale, NJ: Erlbaum. • Laio, 2008. A defense of intuitions. Philosophical Studies, 140:2, 247-262. (methodological and conceptual difficulties of a conclusion that due to variability intuitions are not reliable.) • Davis-Floyd, Arvidson (eds). Intuition: The Inside Story. New York: Routledge. • Macnamara 1991. The development of moral reasoning and the foundation of geometry. • Luper (ed.). The Skeptics. • Parsons 1986. Intuition in constructive mathematics. In J. Butterfield (ed.). Language, Mind, and Logic (pp. 211-229). Cambridge: Cambridge University Press. • Wright 2004. Intuition, entitlement, and the epistemology of logical laws. Dialectica, 58:1, 155-175.

● ● ● Nagel, J. (2012). Intuitions and Experiments: A Defense of the Case Method in Epistemology. *Philosophy and Phenomenological Research*, *85*(3), 495–527. doi:10.1111/j.1933-1592.2012.00634.x

Info:

To be dialectically effective, an intuition doesn't have to be correct: it just needs to be shared by one's audience. The argument:

1. Mindreading is responsible for knowledge ascriptions. Mindreading is a good tool because • it's used every day, • it's calibrated against its predictions, • it's universal. Thus, philosophers just use this everyday tool on philosophical cases. There can't be anything wrong with it.
2. Intuitions work according to Self-Consistency Model. Making an intuitive judgment is a stochastic process and thus sometimes may go wrong. SCM gives back both an answer to a question *is X a case of P?* and the confidence level. The confidence level predicts the stability and universality of the answer.
3. However, we have those strange experiments. We can explain the results away by: • subjects' lack of motivation to assess the cases, • under-description of the cases, • errors the stochastic processing in SCM.
4. An analogy to perceptual illusions: order effects like perceiving colors in different contrasts—we see colors differently in different surroundings but it doesn't undermine the reliability of perception. It's more complicated than *a light-off* situation.

Detailed:

1. Intuitive processes update beliefs without an individual's attending to what justifies the modification. *We are conscious only of the result of the computation, not the process* (p. 4). Reflective judgment is explicit. Intuitive judgments may be steps of a reflective judgment (adding huge numbers digit-by-digit). *The fact that there are various differences between intuitive and reflective judgments does not entail that there are any differences in the qualities that make the products of either type of judgment count as knowledge* (p. 6).
2. The abilities used for philosophical thought experiments are used in real-life situations.
3. There exist pre-theoretical intuitions. In the case of some of them, people who experience them • anticipate that in the future they will make the same judgment and • expect others to have the same intuitions. Question: why?
4. When an intuition is strong it's likely to be stable and universal because the strength of an intuition is determined by how easy it is to make the judgment.
5. Koriat's **Self-Consistency Model**. A person's confidence predicts the stability, universality, and validity of an intuition. The confidence is a byproduct of the process of intuitive judgment. The procedure for a binary choice: draw samples of representation from a pool until you reach a threshold or run out of representations. A representation is a consideration for either of the choices. It's stochastic so one can make a choice in favor of one choice when the whole pool overwhelmingly favors the other one. The more homogenous the sample is, the more confident an individual is and the more stable and universal the answer is.
6. *The fact that philosophers agree on Gettier's cases is evidence that we are drawing from similar pools* (p. 15).
7. **Mindreading** is responsible for desires, beliefs, and knowledge attributions. Mindreading makes predictions about the world (how people will behave) and thus individuals can learn from failures of those predictions. The mindreading capacity and its natural illusions are cross-culturally shared.
8. Psychologists classify knowledge and belief as mental states and if anything then belief is harder to attribute (p. 18).
9. *If we take the standard Western responses as normative* (p. 19)—that is as the right ones? *The non-normative responses lie closer to the 50-50 split*—did she have the original data? Anyway, the difference between Ws and EAs might have been due to EAs being not interested in majoring in non-humanities, as they usually do.
10. Her experiment—ordinary knowledge (72%): the clock says 4:15pm and it is 4:15pm; skeptical pressure (40%): the clock is working but Wanda looks at it for too short to be able to tell; Gettier case (33%): the clock is broken but by coincidence, it says the right time; JFB (16%): the clock says 4:53pm but it's 4:15pm. No ethnicity and gender differences. Interpretation: *If undergraduates with little or no philosophical training generally classify epistemology cases the way epistemologists would, and if there is no appre-*

ciable gender or ethnic variation in case responses, then there is no good reason to doubt the default hypothesis that we are relying on a common mindreading capacity in responding to these cases (p. 21).

11. *Amateur participants who do not have the motivation of testing an epistemological theory may be less inclined to read closely* (p. 22).
12. The TrueTemp (Swain et al. 2008) analogy in perception—identifying colors in contrast or in spectrum-lines.
13. Simon Cullen (2009): *consider cases independently*—and the order effect in TrueTemp disappears.

Notable references:

• Koriat, A. (1975) Phonetic symbolism and feeling of knowing. *Memory & Cognition*, 3(5), 545-548. • (1976). Another look at the relationship between phonetic symbolism and the feeling of knowing. *Memory & Cognition*, 4(3), 244-248. • (1995) Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, 124(3), 311-333. • (2008) Subjective confidence in one's answers: The consensuality principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 945-959. • (2011). Subjective Confidence in Perceptual Judgments: A Test of the Self-Consistency Model. *Journal of Experimental Psychology: General*, 140(1), 117-139. • Koriat, A., & Adiv, S. (in press). The construction of attitudinal judgments: Evidence from attitude certainty and response latency. *Social Cognition*. • Kelley, Linsday (1993), Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions, *Journal of Memory and Language* (Vol. 32, pp. 1). • Alter, Oppenheimer (2009), Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13(3), 219. • Reber, Schwarz (1999), Effects of perceptual fluency on judgments of truth, *Consciousness and cognition* (Vol. 8, pp. 338-342). Orlando, Fla.: Academic Press. • Brewer, Sampaio (2006), Processes leading to confidence and accuracy in sentence recognition: A metamemory approach. *Memory*, 14(5), 540-552. • Simmons, Nelson (2006), Intuitive confidence: Choosing between intuitive and nonintuitive alternatives. *Journal of Experimental Psychology-General*, 135(3), 409-427.

● ● ● Schwitzgebel, E., Cushman, F. (2012). Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers. *Mind & Language*, *27*(2), 135–153. doi:10.1111/j.1468-0017.2012.01438.x

Info:

Principle of **moral luck**—we can be morally assessable for outcomes partly outside our control. The difference between **action and omission**—i.e. killing vs. letting die. The doctrine of the **double effect**—it is worse to harm a person as a means of saving others than to harm a person as a side-effect of saving others.

Non-philosophers, philosophers, and PhDs in ethics are influenced by the order of cases, i.e. 54% of philosophers say that pushing in the trolley problem is not morally worse than switching when pushing was presented first, and 73%, when pushing was first. Among PhDs in ethics, respectively, 50% and 75%.

People are influenced by order when claiming to endorse particular moral principles. Thus, moral reasoning seems to be a post-hoc rationalization of non-reflective moral choices which themselves are influenced by non-moral factors.

Notable references:

• moral judgments influenced by the presence or absence of direct physical contact, Cushman, F. A., Young, L., & Hauser, M. D. 2006: *The role of conscious reasoning and intuitions in moral judgment: Testing three principles of harm*. Psychological Science, 17, 1082-1089; • the order in which hypothetical moral scenarios are presented, Lombrozo, T. 2009: *The role of moral commitments in moral judgment*. Cognitive Science, 33, 273-286

● ● ● Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., & Ditto, P. H. (2009). The Motivated Use of Moral Principles. *Judgment and Decision Making*, *4*(6), 476–491.

Info:

Defending moral judgments by appealing to one's subjective preferences is unpersuasive—instead, people appeal to moral principles in justifying their judgments.

Consequentialism = acts are morally right or wrong depending on the net value of their consequences. Deontology = at least some acts are wrong no matter what consequences they have.

Even when utilizing scenarios that have been shown to reliably elicit consequentialist or deontological intuitions, people's moral judgments are often affected by a set of other motivations, such as the desire to protect their ideological beliefs. Many political conservatives, for example, have staked out a clear deontological position in their moral evaluation of embryonic stem cell research, arguing that the potential lives saved by this research do not justify the sacrificing of embryonic life. In their moral assessments of the extensive civilian death toll caused by the invasion of Iraq, however, conservatives have been more consequentialist in tone, arguing that civilian casualties are a necessary cost to achieve a greater good.

Experiment 1. As a pre-study, they checked what people treat as relevant factors for moral judgments; what's important, 87% don't treat race as relevant. In the actual experiment, they presented participants with a trolley problem. In half the scenarios Tyrone Payton could be sacrificed to save 100 members of the New York Philharmonics, in the other half Chip Ellsworth III could be sacrificed to save 100 members of the Harlem Jazz Orchestra. On a Likert scale, they asked whether sacrificing is justified, moral, whether it's sometimes necessary to allow an innocent person to die to save a larger number of innocent people, whether one should never violate certain core principles; from those answers, they combined an index indicating how a strong consequentialist the participant is.

Results: liberals were less willing to endorse the killing of an innocent person on consequentialist grounds when the name of the individual suggested he was Black than when it suggested he was White. No such result for conservatives.

Experiment 2. The same as above, but they gave both stories (Chip & Tyrone), varying order. Results: liberals were more consequentialist when reading about Chip first than in the case of Tyrone. They used the same rule when answering the second question (ϱ=.98; Tyrone (β=-.45)→Chip (-.51), Chip (.19)→Tyrone (.19). Conservatives opposite pattern, but not significant (.5→.26, -.18→ -.1).

Experiment 3. Two versions (Iraqi victims/American victims) of a story: [American/Iraqi] leaders deciding to carry out an attack to stop key leaders of the [Iraqi insurgence/American military] in order to prevent future deaths of [Iraqi insurgents/American troops]. While the decision-makers were aware of the possibility of innocent deaths, they reasoned that sometimes it is necessary to sacrifice innocent people for the sake of a greater good (in this case the saving of many future lives). The decision-makers did not intend the death of any innocent civilians but merely foresaw it as an unwanted consequence of their military actions.

Results: conservatives more likely to endorse consequentialist military action when the victims were Iraqi than when the victims were American. No significance for liberals.

Experiment 4. The same stories. Participants were primed with a patriotic or a multicultural condition: a task of unscrambling 11 sentences, 6 neutral and (a) 5 with words *patriots*, *USA*, *flag*, *loyal*, (b) 5 with words *multicultural*, *include*, *diversity*, *minority*. Results: For the Iraqi scenario, participants primed with patriotism were more likely to endorse consequentialist military action than were participants primed with multiculturalism. No significance for American victims scenario.

Notable references:

• Aiken, L. S. & West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Sage Publications, Inc; London. • Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype priming on the action. *Journal of Personality and Social Psychology*, 71, 230–244. • Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgments and decision-making. *Cognition*, 108, 381–417. • Bartels, D. M., & Medin, D. L. (2007). Are morally motivated decision-makers insensitive to the consequences of their choices? *Psychological Science*, 18, 24–28. • Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: The use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, 63, 568–584. • Dunning, Leuenberger, Sherman (1995). A new look at motivated inference: Are self-serving theories of success a product of motivational forces? *Journal of*

Personality and Social Psychology, 69, 58–68. • Haidt (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814–834. • Haidt, Graham (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20, 98–116. • Haidt, Rosenberg, Ho (2003). Differentiating diversities: Moral diversity is not like other kinds. *Journal of Applied Social Psychology*, 33, 1–36. • Jost, Glaser, Kruglanski, Sulloway (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, 129, 339–375. • Paharia, Deshpandé (2009). Sweatshop labor is wrong unless the jeans are cute: Motivated moral disengagement. *Harvard Business School Working Paper*, No. 09–079, January 2009. • Tetlock, Kristel, Elson, Green, Lerner (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, 78, 853–870.

Framing Effects

● ● ● Tversky, A, Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, *211*(4481), 453–458. doi:10.1126/science.7455683

Info:

(Economically) rational choices satisfy basic requirements of consistency and coherence. A decision problem is defined by the acts to choose, the possible outcomes, and the conditional probabilities that relate outcomes to acts. A decision frame = the decision-maker's conception of the acts, outcomes, and contingencies associated with a particular decision problem. The frame is controlled both by the formulation of the problem and by the norms, habits, and personal characteristics of the decision-maker.

The vignettes. **Problem 1** [N = 1521: Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

If Program A is adopted, 200 people will be saved. [72%]

If Program B is adopted, there is a 1/3 probability that 600 people will be saved and a 2/3 probability that no people will be saved. [28%

Problem 2 [N = 155]: If Program C is adopted 400 people will die. [22%]

If Program D is adopted there is a 1/3 probability that nobody will die, and a 2/3 probability that 600 people will die. [78%]

Choices involving gains are often risk-averse and choices involving losses are often risk-taking. However, the two problems are effectively identical.

The **prospect theory**. ν – evaluation of outcomes, π – weights of probabilities; the overall value of a problem $\{(x, p), (y, q)\}$: $\pi(p)\,\nu(x) + \pi(q)\,\nu(y)$. Outcomes are evaluated as positive (gains) or negative (losses) deviations from a neutral reference outcome $\nu=0$. π is monotonic over probabilities but starts above *id* and quickly goes a bit below; low probabilities are overweighted and moderate and high probabilities underweighted.

Some other interesting problems:

Problem 5 [N = 77]: Which of the following options do you prefer?

A. a sure win of $30 [78%]

B. 80% chance to win $45 [22%]

Problem 6 [N = 85]: Consider the following two-stage game. In the first stage, there is a 75% chance to end the game without winning anything, and a 25% chance to move into the second stage. If you reach the second stage you have a choice between:

C. a sure win of $30 [74%]

D. 80% chance to win $45 [26%]

Your choice must be made before the game starts, i.e., before the outcome of the first stage is known. Please indicate the option you prefer.

Problem 7 [N = 81]: Which of the following options do you prefer?

E. 25% chance to win $30 [42%]

F. 20% chance to win $45 [58%]

If the second stage of the game is reached, then problem 6 reduces to problem 5; if the game ends at the first stage, the decision does not affect the outcome. Hence there seems to be no reason to make a different choice in problems 5 and 6. Problem 6 is equivalent to problem 7 on the one hand and problem 5 on the other. It is not obvious which preferences should be abandoned.

Individuals who face a decision problem and have a definite preference (i) might have a different preference in a different framing of the same problem, (ii) are normally unaware of alternative frames and their potential effects on the relative attractiveness of options, (iii) would wish their preferences to be independent of frame, but (iv) are often uncertain how to resolve detected inconsistencies. Preference reversals, or other errors of choice or judgment, don't have to be irrational. The practice of acting on the most readily available frame can sometimes be justified by reference to the mental effort required to explore alternative frames and avoid potential inconsistencies (*bounded rationality*).

Notable references:

• J. Von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*. Princeton Univ. Press, Princeton, N.J., 1947. • D. Kahneman and A. Tversky, *Econometrics* 47, 263 (1979). (A detailed description of the prospect theory). • Simon, *Q. J. Econ.* 69, 99 (1955); Psychol. Rev. 63, 129 (1956). (*bounded rationality*).

● ● ● Druckman, J. N. (2001). Evaluating Framing Effects. *Journal of Economic Psychology*, *22*(1), 91–101. doi:10.1016/s0167-4870(00)00032-5

Info:

After Tversky and Kahneman, analogous experiments were conducted but their results weren't of smaller magnitude. This one replicated the original results. Druckman added one more vignette with the output described both as gains and as losses (i.e. *200 people will be saved and 400 people will die*.) 68% of the participants preferred the risk-aversive alternative in case of *gains* (… *will be saved*), 44% in case of *both gains and losses* (… *will be saved and* … *will die*) and 23% in case of only *losses* (… *will die*).

The additional vignette provides a baseline for evaluating the impact of the frames. Both one-perspective frames were equally distant from the baseline (*gains* +24%, *losses* -21%). Thus, neither the first nor second frame was more influential. In the absence of influence of a format (under the assumption that in the new vignette there is no such influence) participants are minimally risk-seeking.

Notable references:

• Zaller J. (1992). *The nature and origins of mass opinion*. New York: Cambridge University Press. • Levin I., Schneider S., Gaeth G. (1998). All frames are not created equal: a typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Making Process*, 76, 149-188.

● ● ● Frisch, D. (1993). Reasons for Framing Effects. *Organizational Behavior and Human Decision Processes*, *54*(3), 399–429. doi:10.1006/obhd.1993.1017

Info:

Strict framing effect: a set of problems involving redescription of the exact same situation. Loose framing effect: a set of problems equivalent form the perspective of economics. Framing effects violate the principle of *description invariance*—the way a situation is described shouldn't affect one's decision. Description invariance hasn't been criticized as a normative principle.

Decision preference = the choice a person makes; experience preference = the option the person would experience as more desirable.

Two explanations for framing effects: (1) they are like a perceptual illusion—the decision preference is affected by framing, experienced preference is not. People don't reveal their true preferences with their decisions.

(2) framing effects affect the experience of the consequences. In this case, it might be rational for framing to affect choices.

Experiment 1. If a person responds differently to two frames but agrees that they are equivalent, the decision is affected by irrelevant aspects of the description. If a person provides justifications for why the frames actually describe different situations, it is an experience-preference difference. Questions: do subjects who demonstrate framing effects, on reflection agree that the two situations should be treated the same? If not, how do they justify treating them differently?

A set of vignettes, including the original Tversky-&-Kahneman one (no. 7), was used. If the subjects' inferences went beyond the information given in the descriptions, their answers were excluded. The justifications supported the experience preference difference if the subject made reference to regret, fairness, wastefulness, and so on. Two independent judges coded the responses.

In the case of the Asian disease, 69% of the subject who treated the two versions differently stated that they should be treated the same. The ones who said the two problems should be treated differently typically didn't give a clear explanation of their decision. In general, across all of the problems, the rate of changing mind to *identical* was surprisingly low.

Experiment 2. Not that interesting. They looked for consistency of framing effects within categories of problems but didn't find any. Although in general, in the case of Loss/Gain-problems 47% of the subjects changed their mind to *identical*, in the case of Sunk Cost-problems 11% changed their mind.

Conclusions: framing effects occur because (a) the framing affects responses against experience preferences, (b) framing changes experience, (c) for loose framing effects non-economic factors interfere. Usually, most subjects who show framing effects do not agree that the two versions are equivalent even when they directly compare the two versions. It might be because they don't recognize the equivalence even when they see both versions together, or because they feel compelled to justify their original answer.

Notable references:

• Levin, Gaeth, (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *The Journal of Consumer Research*, 15, 374-378. (Subjects' rating of the experience of eating a hamburger is affected by whether the meat is described as *75% lean* or *25% fat*.) • Hoch, Ha, (1986). Consumer learning: advertising and the ambiguity of product experience. *The Journal of Consumer Research*, 13, 221-223. (Advertising affects the evaluation of a product but only when the experience is ambiguous). • Shafer (1986). Savage revisited. *Statistical Science*, 1, 463-501. (*A utility is a value deliberately attached to a consequence created at a given level of description. The consequence is a product of our imagination. The utility is a product of our will*. p. 481). • Tversky, Sattah, Slovic (1988). Contingent weighting in judgment and choice. *Psychological Review*, 95, 371-384. (*But if different elicitation procedures produce different orderings of options, how can preferences and values be defined? And in what sense do they exist?* p. 383).

● ● ● McElroy, T., Seta, J. J. (2003). Framing Effects: An Analytic–Holistic Perspective. *Journal of Experimental Social Psychology*, *39*(6), 0–617. doi:10.1016/s0022-1031(03)00036-2

Info:

Framing effect = when a decision-maker responds differently to different but objectively equivalent descriptions of the same problem. The decision-making process is dichotomous, consisting of two phases: (a) editing—coding as gains or losses, relative to a neutral reference point—the options, (b) evaluation of the options using weighted probabilities.

The question: **under what conditions** and **for who** are framing effects likely to occur?

Dual-process theories: there are two different processing routes. System 1: automatic and holistic and leads to an automatic contextualization of problems. It relies on contextual cues, uses internal representations of the problem, and allows inferring without detailed scrutiny of the material. It should be sensitive to contextual cues, such as how the problem is framed. System 2: controlled and analytic processing style and serves to decontextualize and depersonalize problems. It abstracts problems into canonical representations devoid of context. Individuals with a high level of analytic intelligence more likely to engage in system 2 processing;

each individual's style of epistemic regulation may be relatively consistent. Holistic/heuristic processing normally occurs when an individual has low levels of motivation or ability. Analytic/systematic processing is significantly more effortful, it only occurs when individuals are both willing and able to perform the task.

Experiment 1. The Asian disease case, once as personally relevant, once as irrelevant. Assumption: in the first case participants use System 1, in the second one, System 2. Results: Framing effect occurred only in the personally irrelevant condition.

Experiment 2. Participants were examined for styles of thinking using the Zenhausern preference test (20 questions including *are you logical?*). The gain/loss frame had a significant impact on System 1 participants and a marginal impact on System 2 participants.

Conclusions: Holistic (system 1) processing is prone to framing; analytic (system 2) processing is immune to framing. Framing effects are not observed when the decision has consequences to the decision-maker.

Notable references:

• Stanovich, West, Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23, 645–726. (2000) (They've used the dual-process theories of reasoning to explain why some individuals behave in a maximizing rational way whereas others do not). • Epstein, Lipson, Holstein, Huh (1992). Irrational reactions to negative outcomes: Evidence for two conceptual systems. *Journal of Personality and Social Psychology*, 38, 889–906. • Evans, Over (1996). *Rationality and reasoning*. Psychology Press. • Sloman (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3–22. • Chaiken (1987). The heuristic model of persuasion. In Zanna, Olson, Herman (Eds.), *Social influence: The Ontario Symposium*: Vol. 5 (pp. 3–39). Hillsdale, NJ: Erlbaum • Fiske & Neuberg, 1990, • Petty, Cacioppo (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer. (how contextual and cognitive style differences influence an individual's processing style and ultimately social judgment). • Zenhausern (1978) Imagery, cerebral dominance, and style of thinking: A unified field model. *Bulletin of the Psychonomic Society*, 12, 381–384. (preference test (PT). The PT is an index of cognitive style consisting of 20 items). • O'Connor, Pennie, Dales (1996). Framing effects on expectation, decisions, and side effects experienced. *Journal of Clinical Epidemiology*, 49, 1271–1276 (for patients who were actually in the position of receiving a vaccination injection, decisions about receiving the vaccine were not affected by framing).

● ● ● Armstrong, K., Schwartz, J. S., Fitzgerald, G., Putt, M., Ubel, P. A. (2002). Effect of Framing as Gain versus Loss on Understanding and Hypothetical Treatment Choices: Survival and Mortality Curves. *Medical Decision Making*, *22*(1), 76–83. doi:10.1177/0272989X0202200108

Info:

For each treatment option, the survival curve shows the percentage of people living each year after being diagnosed, mortality curve—the percentage of people who die. The participants were presented with survival curves, mortality curves, or both. Participants who received the information framed as both survival and mortality curves were less likely to choose preventive surgery than participants. receiving survival curves only and more likely to choose preventive surgery than participants receiving mortality curves only.

What might be important is that when presented with mortality curves alone, the participants did worse answering control questions (i.e. how many people are still alive, how many had already died) than in other cases. Similar control questions weren't asked in Tversky & Kahneman's or Druckman's experiment.

Notable references:

• Tversky A, Kahneman D. Loss aversion in riskless choice: a reference dependent model. Q J Econ. 1991; 107:1039–61. • Miller GA. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol Rev. 1956; 63:81–97.

● ● ● Andow, J. (2017). Are Intuitions About Moral Relevance Susceptible to Framing Effects? *Review of Philosophy and Psychology*, *9*, 115–141. doi:10.1007/s13164-017-0352-5

● ● ● Weinberg, J. M., Alexander, J., Gonnerman, C., Reuter, S. (2012). Restrictionism and Reflection. *Monist*, *95*(2), 200–222. doi:10.5840/monist201295212

Info:

The experiment. The story: Imagine that… supercomputer which can deduce…exactly what will be happening. The computer deduces that Jeremy will definitely rob Fidelity Bank at 6 pm on… The extent to which you agree (min: 1) or disagree (max: 5): • Jeremy could have chosen not to rob the bank, • when Jeremy robs the bank, he acts of his own free will.

The stories in two different fonts: *easy* to read—Arial and *hard*—Mistral.

Results. Jeremy could have chosen not to rob the bank: $\mu_{easy} = 2.37 < \mu_{hard} = 2.70$.

Jeremy acted of his own free will: $\mu_{easy} = 2.08 < \mu_{hard} = 2.46$.

They used the cognitive reflection test (CRT: in a lake, there are lily pads and they double each day…).

Among high-CRT participants the difference was still significant: *easy* (2.17, 2.00) < *hard* (2.96, 2.56).

The explanation. Participants answer the question *how easy would it be for Jeremy to do otherwise?* based on how easy it is for them to read the story. So, it's hard for me to read the story → it's hard for Jeremy to do otherwise. What matters is the experience of fluency while reading. And that doesn't have to be achieved only with fonts—it might be visual, linguistic, conceptual.

Notable references:

• Cushman and Schwitzgebel, forthcoming, • Haird, Koller, Dias, 1993, • Nichols, Alexander, Weinberg, 2008, • Nichols, Knobe, 2007, • Petrinovich, O'Neill, 1996, • Swain, Alexander, Weinberg, 2008, • Weinberg, Alexander, Gonnerman, Reuter, forthcoming. • Alter and Oppenheimer, 2009.

● ● ● Valdesolo, P., DeSteno, D. (2006). Manipulations of Emotional Context Shape Moral Judgment. *Psychological Science*, *17*(6), 476–477. doi:10.1111/j.1467-9280.2006.01731.x

Info:

moral judgments are often mediated by two classes of brain processes—(a) processes that automatically alter hedonic states in response to specific types of socially relevant stimuli, and (b) more domain-general, effortful processes that underlie abilities for abstract reasoning, simulation, and cognitive control. Often, these work in unison to foster decisions in accord with the goals of both; goals that are socially adaptive are often congruent with more abstract moral principles. There are dilemmas, however, where those two compete—i.e. dilemmas where one endorses a personal moral violation in order to uphold a utilitarian principle.

I.e. in the footbridge dilemma, which is thus structured, people think it's wrong to push the guy even though it saves people's lives. The rare answer 'push him' is associated with activation of deliberative centers aimed at cognitive control, suggesting that the automatic negative reaction must be disregarded to choose the utilitarian option.

If affective states are a basis for moral judgments for the automatic processes, by manipulating environmental factors separate from any moral violations might influence affect at the time of judgment and thus the judgment itself.

Experiment: participants received a positive (*Saturday Night Live*) or neutral (*Spanish village*) affect induction and immediately afterward were presented with the footbridge and trolley dilemmas in random order, embedded in a small set of non-moral distractors. The trolley dilemma is logically equivalent to the footbridge dilemma, but does not require consideration of an emotion-evoking personal violation to reach a utilitarian outcome; consequently, the vast majority of individuals select the utilitarian option for this dilemma. Each dilemma was presented through a series of three screens, the first two explaining the dilemma and the last asking the participant to indicate whether a described course of action would be *appropriate* or *inappropriate*.

Results: after *Saturday Night Life* people reported a more positive affective state than after the Spanish village (μ=4.57 vs. μ=2.77). In the control group, 3 individuals said the utilitarian response for the trolley dilemma

was appropriate and 35 that it wasn't. In the positive-affect group 10 said it was appropriate and 31 that it wasn't. Longer decisions time increased the probability of choosing the *appropriate* response.

Notable references:

All following for two classes of processes underlying moral judgments: fast & analytic: • Greene, Nystrom, Engell, Darley, Cohen (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44, 389–400. • Greene, Sommerville, Nystrom, Darley, Cohen (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, 2105–2108. • Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814–834.

● ● ● DeSteno, D., Dasgupta, N., Bartlett, M. Y., Cajdric, A. (2004). Prejudice From Thin Air. The Effect of Emotion on Automatic Intergroup Attitudes. *Psychological Science*, *15*(5), 319–324. doi:10.1111/j.0956-7976.2004.00676.x

Info:

Theoretical basis: a functional view of emotions as phenomena designed to increase adaptive responses to environmentally significant stimuli. To the extent that outgroups often signify sources of conflict, competition, or blockage of goals, and to the extent that emotions help individuals meet environmental challenges by activating goal-driven action tendencies, emotions that prepare organisms to meet challenges related to conflict or competition (e.g., anger) should bias automatic intergroup evaluations in accord with these functional goals.

Automatic beliefs and attitudes toward groups are not as immutable as previously theorized, but rather are quite sensitive to external cues such as social context. Just as anger can originate from current interactions with groups, so may incidental feelings of anger from an unrelated situation affect automatic appraisals of social groups. Incidental feelings of anger are likely to increase automatic bias against an outgroup because anger increases negativity toward the outgroup, decreases positivity, or both.

Experiment 1. Pretty complex method. Two groups were assembled randomly but the participants were said they are divided based on underestimating or overestimating (of # of MTA passengers). In the end, angry participants were slower to associate positive attributes than negative attributes with the outgroup. There was no difference in the speed with which they associated positive versus negative attributes with the ingroup. Experiment 2. When participants were angry, reaction time for evaluating ingroup as good/outgroup as bad was lower than for outgroup as good and ingroup as bad. When neutral or sad, the effect didn't occur or was weaker.

Notable references:

• Sadness promotes systematic processing of information that decreases stereotypic judgments / if people suspect that incidental emotion may unduly influence an unrelated judgment, they often correct for the perceived bias: Lambert, Khan, Lickel, Fricke, (1997). Mood and the correction of positive versus negative stereotypes. *Journal of Personality and Social Psychology*, 72, 1002–1016. • if people suspect that incidental emotion may unduly influence an unrelated judgment, they often correct for the perceived bias: DeSteno, Petty, Wegener, Rucker (2000). Beyond valence in the perception of likelihood: The role of emotion-specificity. *Journal of Personality and Social Psychology*, 78, 397–416. • happy individuals, who typically engage in heuristic processing, are able to process systematically when instructed to do so Queller, S., Mackie, D.M., & Stroessner, S.J. (1996). Ameliorating some negative effects of positive mood: Encouraging happy people to perceive intragroup variability. *Journal of Experimental Social Psychology*, 32, 361–386. • or when counter-stereotypic information motivates them to do so: Bless, Schwarz, Wieland (1996). Mood and the impact of category membership and individuating information. *European Journal of Social Psychology*, 26, 935–959. • Such control, however, is not available for all types of judgments, especially automatic ones: Banaji, Dasgupta (1998). The consciousness of social beliefs: A program of research on stereotyping and prejudice. In Yzerbyt, Lories, Dardenne (Eds.), *Metacognition: Cognitive and social dimensions* (pp. 157–170). Thousand Oaks, CA: Sage. • a functional view of emotions as phenomena designed to increase adaptive responses to environmentally significant stimuli: Damasio (1994). *Descartes' error*. New York: Avon Books.; Keltner, Gross, (1999). Functional accounts of emotion. *Cognition and*

Emotion, 13, 467–480; LeDoux (1996). *The emotional brain*. New York: Simon & Schuster. • automatic beliefs and attitudes toward groups are not as immutable as previously theorized, but rather are quite sensitive to external cues such as social context: Dasgupta, Greenwald (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, 81, 800–814.; Wittenbrink, Judd., Park (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, 81, 815–827.

● ● ● Bartlett, M. Y., DeSteno, D. (2006). Gratitude and Prosocial Behavior: Helping When It Costs You. *Psychological Science*, *17*(4), 319–325. doi:10.1111/j.1467-9280.2006.01705.x

Info:

Gratitude = the positive emotion one feels when another person has intentionally given or attempted to give, one something of value. Gratitude functions to nurture social relationships through its encouragement of reciprocal, prosocial behavior between a benefactor and recipient.

The commitment problem = individuals must overcome the worry that they will expend time and resources building a relationship only to receive little or nothing in return. For instance, when deciding whether to enter into a social exchange or economic partnership, one must determine how likely the other person is to uphold his or her end of the bargain. Emotions, such as gratitude, guilt, and love, may play a pivotal role in building trust by encouraging one to adopt behaviors that support the partnership even when such behaviors are costly to oneself in the short term.

From a functionalist view, emotions motivate individuals to behave in ways that help them solve challenges of adaptive import; they should help guide decisions about social exchange in a social species. Gratitude functions to encourage an individual to reciprocate a favor, even if such reciprocation will be costly to him or her in the short term. Over time, this reciprocal prosocial behavior aids in building trust and, consequently, preserving relationships.

One should distinguish the effect of gratitude from that of the reciprocity norm (i.e., cognitive awareness that one should repay another person who has provided assistance). Under certain circumstances, gratitude can facilitate prosocial behavior in a way that a social norm isolated from emotional reactions cannot.

Study 1. The aim—to demonstrate gratitude's direct effect on costly helping behavior and to differentiate this effect from the influence of simple positivity and awareness of reciprocity constraints. The procedure: paired, a confederate blind to the hypothesis. Three conditions: gratitude—the confederate helps with to find out why the screen went off, amusement—*Saturday Night Live*, neutral—nothing. Tests to check whether appropriate emotions were elicited. After the study, the confederate asks the participant to fill out a long survey containing demanding tasks. The time of filling out the survey is the measure of the outcome. A meditational analysis was used to hold the reciprocity constant and measure only the effect of gratitude.

Results: the awareness of receiving help had no additional casual effect on helping behavior beyond the effect triggered by gratitude—so gratitude fully mediated between receiving help and helping. The effect of gratitude (μ=29.94) was higher than for the neutral condition (μ=14.49), which was higher than for amusement (μ=12.11).

Study 2. The aim—to distinguish between gratitude and the effect of being indebted to the partner (for her help). If gratitude triggers a helping behavior toward strangers, it would mean it alone can be a cause of helping behavior. The experiment as in Study 1 but without the amusement condition and in half of the cases a stranger asked for help instead of the helping participant. Results: Participants in the gratitude condition helped more than those in the neutral condition regardless of whether the person asking for help was a stranger or a benefactor. Though, helping the benefactor in the neutral condition was still higher than helping the stranger in the gratitude condition.

Study 3. Making the participants aware of the distinction between the person toward who they felt gratitude and the person who asks for help. As in Study 2, but in some cases, the experimenter asked *was the other participant who figured out what was wrong with your computer?* Results: participants who weren't asked by the experimenter helped the stranger more (μ=19.71) than those in the neutral condition (μ=11.43) and those questioned by

the experimenter (μ=5.88). The last two results weren't significantly different. That means that when people realize that the feeling is external to the situation, they are more likely to become immune to its effect on helping behavior.

Notable references:

• Emotions should help guide decisions about social exchange in a social species: Keltner, Haidt (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion*, 13, 505–521.

● ● ● Valdesolo, P., DeSteno, D. (2007). Moral Hypocrisy: Social Groups and the Flexibility of Virtue. *Psychological Science*, *18*(8), 689–690. doi:10.1111/j.1467-9280.2007.01961.x

Info:

moral hypocrisy = when individuals' evaluations of their own moral transgressions differ substantially from their evaluations of the same transgressions enacted by others. Question: does moral hypocrisy work also in groups? That is, do people evaluate the same actions of members of in-groups better than those of out-groups.

Experiment. Red task—hard (logic problems, 45min), green task—easy (photo hunt, 10min). Condition 1. Subjects could assign either themselves or a future subject to the green task (leaving the red task respectively to the future subject or to themselves), or use a computer which would choose randomly. Then, among other questions, they were asked whether they acted fairly (7-point scale). Condition 2. Subjects witnessed other participants assigning the green task to themselves and were asked how fairly those acted. Condition 3. Subjects witnessed in-group participants assigning the green task to themselves. In-group and out-group were created as in DeSteno, Dasgupta, *Prejudice From Thin Air, The Effect of Emotion on Automatic Intergroup Attitudes*—overestimators & underestimators. Condition 4. Subjects witnessed out-group participants assigning the green task to themselves.

Results: The subject who acted altruistically or used the randomizer were removed from the analysis. The participants perceived themselves and members of their in-group as more fair (μ~4) than members of the out-group or unaffiliated participants (μ~3). Preservation of a positive self-image appears to trump the use of more objective moral principles.

Moral hypocrisy occurred in emergent groups—new, rather arbitrary, and with no social relations between their members—which means this effect needs nothing but the very fact of belonging to a group.

Notable references:

• *in-group morality* has been posited as a fundamental moral intuition: Haidt, J., & Graham, J. (in press). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*.

● ● ● Valdesolo, P., DeSteno, D. (2008). The Duality of Virtue: Deconstructing the Moral Hypocrite. *Journal of Experimental Social Psychology*, *44*(5), 0–1338. doi:10.1016/j.jesp.2008.03.010

Info:

A dual-process model of moral judgment: an intuitive process + a domain-general, consciously-guided one.

Question: is moral hypocrisy driven by a System 1 process or by a System 2 one? Possibilities are two: (a) hypocrisy could be driven by a discrepancy in automatic intuitions in response to one's own versus another's transgressions. That is, individuals might display an automatic positivity bias for their own transgressions relative to others', with higher-order processes simply functioning to create post hoc justifications for "gut-level" decisions. (b) hypocrisy might be driven by differential activation of higher-order cognitive processes geared toward justification and rationalization of one's own transgressions. That is, although individuals might have negative automatic reactions to both their own and others' transgressions, they may engage in more consciously motivated reasoning when judging their own transgressions in order to maintain a positive self-view.

Experiment. High load introduced to disturb a System 2 process. The rest as in Valdesolo, DeSteno, *Moral Hypocrisy: Social Groups and the Flexibility of Virtue*: the green and red task, and participants choosing between them. Results: without an additional high load task participants judged their own behavior as fairer than the same kind of behavior of other participants. When they had to do an additional task (remembering strings of numbers), their judgments toward themselves and toward other participants were the same.

Conclusion: moral hypocrisy is governed by a dual-process model of moral judgment wherein a prepotent negative reaction to the thought of a fairness transgression operates in tandem with higher-order processes to mediate decision making. When contemplating one's own transgression, motives of rationalization and justification temper the initial negative response and lead to more lenient judgments.

Controlled processing need not always function to "correct" more basic, intuitive responses, but rather can be subject to less admirable motives such as the protection of self-image.

Notable references:

• dual-process model of moral judgment: Cushman, Young, Hauser (2006). The role of reasoning and intuition in moral judgments: Testing three principles of harm. *Psychological Science*, 17(12), 1082–1089. Greene, Nystrom, Engell, Darley, Cohen (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44, 389–400. • Individuals might display an automatic positivity bias for their own transgressions relative to others', with higher-order processes simply functioning to create post hoc justifications for "gut-level" decisions: Haidt, (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814–834. • Humans may have evolved an intuitive aversion to violations of equity, with similar aversions evidenced by certain primate species: Brosnan, de Waal (2003). Monkeys reject unequal pay. *Nature*, 425, 297–299. Hauser (2006). *Moral minds: How nature designed our universal sense of right and wrong*. New York: Harper Collins. • Humans have evolved specific social emotions designed to foster cooperation and trust with others suggesting an important role for emotional responses designed to inhibit self-serving behavior and, thereby, to avoid negative social consequences: Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20, 98–116. • Bandura (1996). Mechanisms of moral disengagement in the exercise of moral agency. Journal of Personality and Social Psychology, 71(2), 364–374. • Cognitive load was manipulated using a digit-string memory task: Gilbert, Hixon (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, 60(4), 509–517.

● ● ● DeSteno, D., Petty, R. E., Rucker, D. D., Wegener, D. T., Braverman, J. (2004). Discrete Emotions and Persuasion: The Role of Emotion-Induced Expectancies. *Journal of Personality and Social Psychology*, *86*(1), 43–56. doi:10.1037/0022-3514.86.1.43

Info:

Expectancy-value model: an attitude toward an object is a function of the values one attaches to the object's attributes weighted by the likelihood of occurrence of those attributes in the object. The primary purpose of the emotions is to engender adaptive responses to distinct situational appraisals. If so, then evaluating likelihoods should be affected by emotions if those emotions are related to attributes of an object. But the emotions must match the attributes to be a cue for increasing perceived probabilities. Though, the persuasion will work only on those who relate the emotions to the attributes which require more effortful processing.

Study 1. The hypothesis: increased persuasion would occur under conditions where receivers' emotional states matched the specific emotional overtones of a message as determined by the emotional consequences of the arguments contained within it. Procedure: presenting individuals experiencing either a sad or a neutral state with one of two equally strong versions of a proposal for an increase in their city's sales tax: sadness framed and anger framed. In both versions, the description of the tax increase was identical. In the sadness-framed version, the increase was described as being necessary to combat a series of saddening problems (e.g., the plight of special-needs infants). In the anger-framed version, a series of angering problems in need of remediation (e.g., increasing traffic delays) was listed. A trick: The experimenter informed them that they would be participating in two separate studies: one designed to examine memory for and opinions about events described in popular media outlets (e.g., magazine articles) and the other to evaluate government poli-

cies under consideration. Actually, the articles were used to induce sadness/anger. Results: sad participants were more likely to indicate that they would vote for the sadness-framed tax proposal than were neutral participants; this pattern was reversed among those who received the mismatched anger-framed tax proposal. The tax was supposed to be increased in the state the participants lived in so they would engage in more effortful thinking.

Study 2. Participants were made to experience either sadness or anger and then presented with either a sadness- or anger-framed tax proposal (so: 2×2). The Need for Cognition was measured. Results: the manipulation worked for high-NC participants; it didn't for low-NC ones. Sad individuals high in NC were more favorable toward the sadness-framed proposal than were their angry counterparts. Angry individuals high in NC were more favorable toward the anger-framed proposal than were their sad counterparts. For low-NC participants, anger led to a greater rejection of both messages than did sadness.

Conclusion: emotions induce biases in expectances. That leads to different evaluations of choices.

Notable references:

• the experience of specific emotional states can influence the perceived likelihoods of events matching these states in emotional overtone: DeSteno, Petty, Wegener, Rucker (2000). Beyond valence in the perception of likelihood: The role of emotion specificity. *Journal of Personality and Social Psychology*, 78, 397–416. • the role played by likelihood estimates, or expectancies, in attitude structure / increased fear can be associated with more positive attitudes toward policies or behaviors designed to remedy the fear-inducing threat as long as certain conditions are met (e.g., providing specific fear-reducing actions that can be taken); for a review: Petty, Wegener (1998a). Attitude change: Multiple roles for persuasion variables. In Gilbert, Fiske, Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 323–390). Boston: McGraw-Hill. • Many models of attitude structure highlight the important roles played not only by the desirability of the attributes or outcomes of an attitude object but also by the likelihoods that the object possesses or will result in these attributes or outcomes: McGuire & McGuire (1991). The content, structure, and operation of thought systems. In Wyer Jr. & Srull (Eds.), *Advances in social cognition* (Vol. 4, pp. 1–78). Hillsdale, NJ. • message-induced changes to the desirabilities or likelihoods associated with an object's attributes can result in corresponding attitude change when individuals devote a relatively high level of effort to thinking: Erlbaum. Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In Gilbert, Fiske, Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 269–322). Boston: McGraw-Hill. • Negative affect inflated and positive effect deflated likelihood estimates for the occurrence of negatively toned events (e.g., contracting cancer, being hit by lightning): Johnson., Tversky (1983). Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology*, 45, 20–31. • The *need for cognition* test: Cacioppo, Petty, (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42, 116–131. Cacioppo, Petty, Kao (1984). The efficient assessment of the need for cognition. *Journal of Personality Assessment*, 48, 306–307. • Although situational variables can have a large impact on individuals low in *Need for Cognition*, individuals high in NC are relatively immune to these situational variations: Cacioppo, Petty, Feinstein, Jarvis (1996). Individual differences in cognitive motivation: The life and times of people varying in need for cognition. *Psychological Bulletin*, 119, 197–253. • Wolf (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage. • Tiedens, Linton (2001). Judgment under emotional certainty and uncertainty: The effects of specific emotions on information processing. *Journal of Personality and Social Psychology*, 81, 973–988.

● ● ● Condon P., DeSteno, D. (2011). Compassion for One Reduces Punishment for Another. *Journal of Experimental Social Psychology*, *47*(3), 0–701. doi:10.1016/j.jesp.2010.11.016

Info:

Hypothesis: compassion can decrease the severity of punishment for a transgressor even if it's not directed toward the transgressor.

Experiment: a participant had two fake co-participants. In the control condition, they just finished a task given by the experimenter. In the *cheater with compassion* and *cheater without compassion* conditions: one of them cheated to get some money (and the participant witnessed that). In the *control* and *cwoc* condition the other co-participant left. In the *cwc* condition, the other co-participant said her brother found out recently he has can-

cer. Thus, they inducted compassion but not toward the cheater. Then, the participant could administrate hot sauce to the cheater (or the first participant in the control condition).

Results: In the control condition and the cheater with compassion condition the participants administered ~2g of hot sauce. In the cheater without compassion condition, they administered ~10g.

Conclusions: compassion reduces the desire to punish even if the feeling is not directed at the transgressor.

Notable references:

• Lieberman, Solomon, Greenberg, McGregor (1999). A hot new way to measure aggression: Hot sauce allocation. *Aggressive Behavior*, 25, 331−348. • Preacher, Hayes (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36, 717−731. • Wilkowski, Robinson, Troop-Gordon (2010). How does cognitive control reduce anger and aggression? The role of conflict monitoring and forgiveness processes. *Journal of Personality and Social Psychology*, 98, 830−840.

In Cognitive Psychology

1. Chen, S., Duckworth, K., Chaiken, S. (1999). Motivated Heuristic and Systematic Processing. *Psychological Inquiry*, *10*(1), 44–49. doi:10.1207/s15327965pli1001_6
2. Chaiken, S. (1987). The Heuristic Model of Persuasion. In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.), *Ontario Symposium on Personality and Social Psychology. Social Influence: The Ontario Symposium, Vol. 5* (p. 3–39). Lawrence Erlbaum Associates, Inc.
3. Evans, J. St. B.T., Over D. E. (1996). *Rationality and Reasoning.* London: Psychology Press.
4. Fiske, S. T. (1990). A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation. *Advances in Experimental Social Psychology*, *23*, 1–74. doi:10.1016/S0065-2601(08)60317-2
5. Guglielmo, S. (2015). Moral Judgment as Information Processing: An Integrative Review. *Frontiers in Psychology*, *6*. doi:10.3389/fpsyg.2015.01637
6. Petty, R. E., Cacioppo, J. T. (1986). *Communication and Persuasion Central and Peripheral Routes to Attitude Change*. New York: Springer-Verlag. doi:10.1007/978-1-4612-4964-1
7. Wyer, Jr., R. S., Srull, T. K. (Eds.). (1990). *Content and Process Specificity in the Effects of Prior Experiences* (Advances in Social Cognition, Volume III). London: Psychology Press.