IN DEFENSE OF WEIRD HYPOTHETICALS

Barbara A. Spellman

Professor of Law; Professor of Psychology
University of Virginia

«Our statements about the external world face the tribunal of sense experience not individually but only as a corporate body».
Willard van Orman Quine [1]

ABSTRACT: Professor Allen (this issue) critiques the value of using "weird" hypotheticals to mine intuitions about legal systems. I respond by supporting the value of "thin" hypotheticals for providing information about how people reason generally, rather than for revealing peoples' specific answers. I note that because legal systems are the products of many minds thinking about how other minds operate, the object of inquiry is metacognition—that is, understanding how reasoning works.

KEYWORDS: legal epistemology; reasoning; metacognition; psychology and law

---

[1] Quine, 1951.

1. INTRODUCTION

It is difficult to be a refutator of a refutator: in his essay[2], Ron Allen exhibits some deft refutating. It is also difficult for me to want to agree with Ron Allen about anything—even when I think he's right[3]. And I do think he's right that relative plausibility is a better explanation (than probabilities) of how people reason about evidence, and, in fact, about many things. More specifically, as a cognitive scientist, I agree with his conclusions: that the attempt to explain evidentiary reasoning with probabilities is misguided and that what I will call «thin hypotheticals» (rather than «weird hypotheticals»)[4] are misused in trying to gather information useful for understanding our legal system. However, in my view, the fact that they can be misused does not mean that they cannot be extremely useful.

In Section 2, I describe why I believe that naturalized epistemology is a good way to think about human reasoning generally. There I lay the groundwork for the main argument in Section 3—that thin hypotheticals are, in fact, quite useful to gain understanding about how people reason about law and evidence. Part of my argument is that I often don't care much about «what» people say in response to those hypotheticals but I care about «why» they say it. Empiricists who believe that people are giving odd or incorrect answers can try to chase down whether it is because of experimenter demands, the way the information is presented, the particular contents of the hypothetical, incompetence with statistics, or some general characteristic of human reasoning. We should care *why* because when we understand the cause, we can get ideas about what other situations would result in similar «bad» answers and how to fix them.

In Section 4, I'll take us for a ride on the blue bus and consider relative plausibility. Finally, Section 5 steps back to consider how the legal system is a complex adaptive system and describes how thinking about how other people do think, and

---

[2] ALLEN, 2020 (this issue).

[3] I once cut off an interminable audience member «question» at the end of a long conference day. Professor Allen took a vote among the attendees regarding my, admittedly, controversial behavior.

[4] Allen uses the term «weird hypotheticals» and argues that they ignore important aspects of the legal system including its values. I believe that the modifier «thin» as introduced by Gilbert Ryle and made popular by the anthropologist Clifford Geertz captures those deficiencies better than «weird». A «thin» description is a surface level one, often of just a behavior; a «thick» description puts a behavior within a context or culture.

can think, (i.e., metacognition) is a more essential foundation of the system than is relative plausibility.

2. A NATURALISTIC TURN: THE RISE AND FALL AND RISE OF HEURISTICS AND BIASES

Let's start with a very brief and somewhat idiosyncratic view of the recent history of the psychology of judgment and decision making[5]. In the 1950's and 1960's[6], there was some psychology research that endorsed the view of the «Rational Man», with studies that were interpreted as showing that humans are good at statistics and other types of formal reasoning.

But things turned in the late 1960's and 1970's; not all was found to be well with human's ability to think logically or statistically. An early important paper by Tversky and Kahneman (1974) called «Judgment under Uncertainty: Heuristics and Biases» demonstrated how experiment participants made various kinds of reasoning errors when presented with simple reasoning tasks that relied on basic concepts from logic or statistics. The authors concluded: «What is perhaps surprising is the failure of people to infer from lifelong experienc […] fundamental statistical rules…».[7] For the next 20 years, during the «Reign of Heuristics and Biases»[8], discovering different ways in which human reasoning was flawed became a cottage industry[9] for experimental psychologists who argued, in effect, that people are stupid. But that was not Tversky and Kahneman's point: they acknowledged that people were making mistakes, but the interesting thing was that the mistakes were systematic rather than random. Thus, these errors could be useful to researchers when trying to uncover what people[10] are really doing when they are, supposedly, or should be, using logic or statistics[11].

The 1990's were filled with explanations for the errors. Some lines of work argued that study participants were getting the «wrong» answer because they were filling in

---

[5] See Spellman and Schnall, 2009: 126-130, for a less brief but equally idiosyncratic explication.

[6] Peterson and Beach, 1967.

[7] Tversky and Kahneman, 1974: 1130.

[8] Spellman and Schnall, 2009: 127

[9] This cottage industry was probably located next door to the cottage industry that Allen mentions (ms p. 3), where unnamed disciplines create weird hypotheticals.

[10] It is likely that all, and certain that most, of the experiments that I describe in this article had participants who were «WEIRD» people. WEIRD is an acronym for Western, Educated, Industrial, Rich, and Democratic—qualities of the people and their countries who comprised the overwhelming majority of psychology research participants up to 2010 (and continue to do so). Henrich, Heine & Norenzayan, 2010.

[11] I am happy to state that I believe that people are «bad» at the conscious use of probabilities and statistics. But I also believe that people often make judgments «as if» they are using probabilities and statistics without even knowing that they do so.

missing information in the sketchy experimental problems or vignettes with other (typically sensible) information [12]. Another line of work argued that evolution could explain the errors, that humans did not evolve using probability theory, and that people were much better at reasoning when numbers are described in terms of frequencies rather than probabilities [13]. And a line of work that grew out of Kahneman's lab suggested that many of the errors arose from «attribute substitution»—that when a problem is missing information, or too difficult, or has to be done under time pressure, reasoners will substitute an easy-to-think about way of answering rather than grappling with the difficult demands [14].

One example, relevant to the blue bus problem, comes from the base-rate fallacy. Experiment participants read brief personality descriptions of people who were said to be either engineers or lawyers. The descriptions were said to be drawn from a group of 100 descriptions, of which 70 were engineers and 30 were lawyers, or vice versa. Participants were remarkably insensitive to the proportions, making about the same profession-probability judgments for each personality description regardless of the base-rate condition.

Another famous example is the «Linda problem», for which many people fall prey to the conjunction fallacy. Participants read about an intelligent 31-year old woman who, as a college student, was concerned with progressive issues and had participated in demonstrations. When asked whether it was more likely that Linda was a bank teller or that Linda was a bank teller who was active in the feminist movement, participants chose the latter, despite that answer being logically impossible [15]. In each of these examples, rather than engaging in complicated thinking about likelihoods, participants were engaged in the simpler task of thinking about what the subjects of the vignettes were like.

Dual process theory developed out of the recognition that there were similar patterns of evidence in a wide range of research domains [16]. It suggests that people use two different paths in reasoning: the «heuristic system» (known as System 1), in which reason is fast and effortless and the «logical system» (known as System 2), in which reason is slow and conscious [17]. Note that although this model provides a convenient way to think about how people reason, and has been useful in getting people to make better decisions, [18] it should be thought of as a metaphor rather than an accurate description of a mechanism.

_____

[12] E.g., CHENG and NOVICK, 1990, for causal attribution
[13] GIGERENZER and HOFFRAGE, 1995.
[14] KAHNEMAN and FREDERICK, 2002.
[15] KAHNEMAN and TVERSKY, 1982.
[16] SMITH and DeCOSTER, 2000.
[17] There are other qualities assigned to the two systems and debates over which are essential qualities. KAHNEMAN, 2011; EVANS & STANOVICH, 2013.
[18] See KAHNEMAN, 2011; THALER & SUNSTEIN, 2009. Of course, it can also be used to find ways to get people to make worse decisions.

3.   WHAT DO «YOU» CARE WHAT OTHER PEOPLE THINK?[19] THE VALUE OF THIN HYPOTHETICALS

I don't care much about what people think of as answers to weird hypotheticals, but I do care why they think it. To figure that out takes a lot of work—more than using the same thin hypothetical over and over to try to tap into «intuitions». To actually learn something, we need at least some comparison conditions, preferably thoughtful ones, and hopefully, other research that reveals the same underlying explanation (or theory) of the result.

The short problems or vignettes from the heuristics and biases era can be viewed as thin hypotheticals and we have learned many things from them. For example, for some problems, most people jump to incorrect intuitive conclusions[20] but others do not; whereas for other problems, the proportions are reversed. What affects how such judgments are made? We know that judgments are affected by characteristics of: the problem—e.g., whether it is phrased using probabilities or frequencies[21]; the person—e.g., people who are high in Need for Cognition are more likely to use the more logical reasoning system[22]; and the environment in which the reasoning takes place—e.g., how much time is available, whether the person is multitasking.

About the blue bus and gatecrasher hypotheticals, Allen states that «for these hypotheticals to be helpful (to jurists), they must be crafted without losing sight of the actual object of inquiry, namely, the legal system»[23]. Well, yes, and no. When I first read Allen's statement, I thought: «But where is the legal system?»—riffing on the philosopher Gilbert Ryle's story of a foreign visitor to Oxford University, who, having been shown colleges, libraries, sports fields, etc., eventually asks: «But where is the university?»[24]. In fact, Allen does provide a very inclusive description of «the American legal system» and juridical proof quoted from Allen and Pardo (2019)[25]. The description includes a lot about how trials are structured and how litigation proceeds using evidence, including «(3) the manner in which humans process and deliberate on evidence».

To my reading, however, the most important part of Allen's description needs us to take a step back to something assumed but not said that encompasses it all: the structure and processes and forms of the systems of proof were designed by human minds to regulate the actions of other humans as they attempt to influence other minds. So, what, really, is the object of inquiry? The legal system or the human

---

[19]   Yes, the title of the 2001 book about Richard Feynman.
[20]   Note that it is not always the case that a fast intuitive answer is wrong whereas the slow conscious one is correct; however, most research problems are set up that way.
[21]   Gigerenzer and Hoffrage, 1995.
[22]   Kahneman and Frederick, 2002.
[23]   Allen, 2020 (this issue): ms p. 272.
[24]   Ryle, 1949: 16.
[25]   Allen, 2020 (this issue): ms pp. 276-277, quoting extensively from Allen & Pardo, 2019: 207-208.

mind (as it interacts with and within that system)? Allen's claim that weird hypotheti-cals ignore important aspects of actual legal systems—because they do not capture all the information in trials, and they tend not to represent all the values that go into a trial decision—is absolutely true. But it is mostly irrelevant.

3.1.   We are all guilty of overusing «simple» tools.

I agree with Allen that we (lawyers, philosophers, psychologists, and others I'm sure) are all guilty of using the same «simple» hypotheticals over and over. Certainly, when a large part of a community knows about the blue bus or about vehicles that manage to sneak into a park,[26] it creates a shorthand way of speaking about ideas to indoctrinated others. But is it useful beyond that?

Thin hypotheticals remind me of thin experiments—or what would be called experiments with low external validity (i.e., unlikely to generalize outside the labora-tory). In experimental psychology, when a simple experimental manipulation creates some interesting and easy-to-obtain data, the field goes wild using it, and often ex-trapolates far beyond its likely relevance.

For example, in the DRM paradigm, experiment participants listen to a list of words (usually nouns) that are all related to one word, which is not presented. In one version, they hear 12 words including: «bed», «rest», «tired», «dream», but they never hear the word «sleep». Immediately after hearing the final word, they are asked to recall the entire list. Typically, about 40% of the participants will recall the never-pre-sented word «sleep». The first paper using this technique[27] has been cited over 4500 times since 1995. The fact that a simple experiment can create a «false memory» is a lovely existence proof that very simple short-term false memories, under conditions of high demand, can be easily created. But how relevant should researchers believe this result is for false memories of witnesses in the courtroom?

Recently the so-called experimental philosophers have been stuck on replay, using the Trolley Problem[28] to study morality so often that it has become a joke and gen-erated dozens of memes. The Trolley Problem seems to me to be an exemplary «thin hypothetical» (and also a «weird hypothetical»)—especially when people are asked «what would you do?» as opposed to «what should a person do»? Researchers can add more people to the tracks, give more time, put the participant in the role of the engi-neer—but what have we learned about real morality? Currently, there is some push-back to contextualize morality research and get away from such thin hypotheticals.[29]

SCHAUER, 2008.
ROEDIGER and MCDERMOTT, 1995.
FOOT, 1978; THOMSON, 1985. The memes can be found by using a Google Image search for «trolley problem».
SCHEIN, 2020.

Using the same thin hypotheticals over and over just to check people's intuitions is certainly not useful. As Allen remarks[30], he often doesn't share the same intuitions as others. Nor should he. Intuitions don't come from nowhere. The heuristics and biases (H&B) literature took advantage of the fact that with thin hypotheticals many or most people jumped to the same wrong intuitive conclusions—but because of personality, knowledge, or training, not everyone jumps. There is a fascinating «adversarial collaboration» paper by Kahneman, whose H&B work often showed intuitions getting people to the wrong answer, and Klein, whose work on naturalistic decision-making more often showed intuitions getting people to the correct answer. They came to a surprising amount of agreement: that there were important differences between the problems they used (Kahneman's hypotheticals vs. Klein's real-world situations); and that intuitions develop when people get reliable feedback about their answers[31]. Of course, not all hypotheticals have a correct answer. Some people pretend that the blue bus hypothetical has a correct answer (liable!) and that anyone who doesn't come to that answer is a fool. But they would be wrong.

3.2. Through Thin and Thick

3.2.1. From Abstract to Concrete to Just Right

Creating variations of hypotheticals that are thicker does not necessarily mean that people will reason better about them. The Wason Selection Task is a deductive reasoning task that was widely used by cognitive psychologists in the 70's and 80's. In the original version, participants are shown four index cards, each with a letter (A or B) or a number (4 or 7) written on it. They are told that each card has a number on one side and a letter on the other. There is a rule that: if there is a vowel on one side, the card must have an odd number on the other side. Participants are then asked to turn over all, and only, the cards they need to turn over to make sure that the rule is enforced. Typically, in this abstract version of the task, most participants turn over the A (to check that there is a 7 on the other side). But few people turn over the other necessary card[32]. Conclusion? People are not good at modus tollens reasoning or falsification? People are not good at abstract reasoning but they would be better if there were context?

In another variation, the four cards had the ages of people in a restaurant on one side—15 and 50—and what they were drinking on the other side—beer or tea. The rule was: if a person is drinking an alcoholic beverage, they must be over 21. In this

---

[30] ALLEN, 2020 (this issue): ms 293-294, near fn 87.
[31] KAHNEMAN and KLEIN, 2009.
[32] That would be the 4—to make sure that there is *not* an A on the other side. WASON, 1968. Wason was a big fan of Popper and claimed that this experiment showed that people were bad at trying to falsify hypotheses.

example, most people correctly turned over the beer card and the 15 card. Why do they get it right? Because it is a culturally familiar situation? Many variations later, we learned that just because the people and activities in the vignettes are familiar, doesn't mean participants would get it right. For example, participants are not so good with the rule: if a person is eating haddock, they must be drinking gin[33]. This is an arbitrary pairing although, of course, people have experience with other food pairing rules.

So, are people only good at deductive reasoning with rules for which they have specific experience? In another set of studies, participants are told about an isolated tribe of people who sometimes wear a bit of volcanic rock. The tribespeople have a rule: if one is going out at night, then one must tie a small piece of volcanic rock around one's ankle. For this version, only 28% of participants knew to check both those going out at night and those not wearing a rock. However, of participants who were told that the tribe believed that vicious spirits roamed the night but a volcanic rock around their ankle would keep them safe, 86% knew who needed to be checked[34]. Thus, people can be good at reasoning about things way outside their experience, and which might seem arbitrary, if they are told that the rule follows a type of culturally common reasoning rule (here, one about precautions). That is, not too abstract, not simply concrete, but in that human reasoning sweet-spot of knowledge-based generalization.

3.2.2.   «Real» Legal Hypotheticals

After this romp through experimental psychology, let's not forget that hypotheticals are also the bailiwick of lawyers and law professors. We use them in classes, at the end of a case excerpts in casebooks, and in legal arguments. Crucially, the hypotheticals we choose tend to be neither thin nor random. They are not thin because they assume a case that is «just like» the current one, about which we are well informed, but with one thing changed. And they are typically not random because the creator of the hypothetical knows what it means to change something that might matter; that is, might make the hypothetical case worthy of a different result. It is in these comparisons that students learn the important features, and the limits, of doctrine; that lawyers argue regarding precedent; and that courts explore possible future implications of their decisions.

Imagine for a second: what would the study of law be without the use of hypotheticals?

---

[33] Manktelow and Evans, 1979. Beer, tea, haddock, and gin: Yes, many of these researchers were British.
[34] Cheng & Holyoak, 1989: 296-299.

4. THE BLUE BUS BEEPS

So back to the blue bus. What are some things that this hypothetical can tell us relevant to the legal system despite it being thin?

Varying the blue bus hypothetical gives us the opportunity to ask questions like: do people answer «not liable» because they don't understand simple probabilities? Are they not sensitive to the fact that 80% is greater than 50%? Do they object to using a 51% threshold for liability judgments? Or are they just answering the question quickly, without really thinking about it, but if forced to engage more deeply they would answer «liable»?

Various experiments based on the blue bus hypothetical have answered those questions—sometimes with «regular people» and sometimes with real judges as experimental participants. For example, Wells[35] asked participants to read variations of the blue bus story and then to estimate «the probability that a bus from the Blue Bus Company killed Mrs. Prob's dog» and to answer the question «If you were a juror in this case, would you rule against the Blue Bus Company and force them to pay damages to Mrs. Prob?» In Experiment 1, psychology student participants read either standard (naked statistics) volume-of-traffic testimony (i.e., that 80% of the traffic was from the Blue Bus Company, which owned 80% of the buses) or «weigh-attendant» (human) testimony (i.e., in which a country transportation official had noted that a blue bus had logged into a nearby weigh station about 10 minutes before the accident and the log book entries were right about bus color 80% of the time). If you've read Allen's essay, you won't be surprised to learn that participants were much more likely to award damages in the weigh-attendant (human) version (67%) than in the standard (naked statistics) version (8%). But for both versions, participants rated the probability of the damage having been done by a blue bus as about 80%.

In Experiment 2, changing the standard statistical version to a rate-of-accidents version (i.e., that 80% of accidents were caused by the Blue Bus Company), didn't change the results of approximately equal probability ratings and much more imposition of liability in the weigh-attendant variation. Experiment 3 added a new condition and also added new types of participants: Master of Business Administration (MBA) students and judges. The patterns looked the same: probability judgments always about 80% and many more plaintiff verdicts in the weigh-attendant than the naked statistics variations.

Another set of blue bus studies[36] found that having participants make a liability decision within 1-2 minutes versus about 6 minutes versus after a group deliberation did not affect probability estimates and did not affect liability decisions. However, in

---

[35] Wells, 1982. Note that in the experiments that follow, in some studies participants were asked both to estimate the probability and whether they would impose liability; in other studies, those questions were asked of different participants.

[36] Wright, Maceachern, Stoffer and Macdonald, 1996.

a version in which the participants read that the Blue Bus Company owned 99.9% of the buses in the area, liability went up from 30% (in the standard condition) to only 63% (in the 99.9% condition).

I learned a lot reading these studies: within these thin hypotheticals, people are sensitive to base rate probabilities, people are more willing to impose liability with greater probabilities but, still, even very very high probabilities alone are not sufficient for them to want to impose liability – an intriguing finding.

My favorite set of blue bus experiments varies how easy it is for participants to construct counterfactuals in which the Grey Bus Company was at fault. In the basic version of the story, participants learned that the bus's tire tracks matched 8 out of 10 of the blue buses and only 2 out of 10 of the grey buses. In the alternative version, other participants learned that the tire tracks partially matched one of the blue buses and one of the grey buses, but the investigator said there was an 80% chance that it was the blue bus. In addition to being asked the probability that a blue bus killed the dog, and whether they would find against the Blue Bus Company, participants were also asked how easy it was to think that a grey bus had run over the dog. As with Wells (1982; which the initial study here was based on), participants did not differ in believing that there was about an 80% chance that it was a blue bus. However, participants varied between conditions in other judgments: those who had been told that the tracks matched 8 blue and 2 grey buses were less likely to find against the Blue Bus Company and said it was easier to think of a grey bus hitting the dog than participants in the 80%-accuracy condition. In fact, participants' ratings of ease of thinking predicted their verdicts. [37]

The finding that the ease of creating counterfactuals affects judgments is not unique to bus accidents. During the Reign of Heuristics and Biases, a busload of studies involved vignettes in which people engaged in «normal» behavior (e.g., taking a typical route) or «unusual-for-them» behavior (which would provoke counterfactual reasoning) [38] and found that people were judged to be more causal of whatever bad outcome occurred to them when they had chosen the abnormal behavior rather than their usual one [39]. Later, various theories of causal attribution looked more precisely at how people used counterfactuals in making causal judgments and found that, among other things, people make such judgments as if they consider the probability of the counterfactual (or counterfactuals) occurring and the conditional probability of its occurrence leading to the outcome in question [40].

---

[37] Niedermeyer, Kerr, and Messé, 1999.

[38] Roese, 1997.

[39] Spellman and Mandel, 1999. The effect of abnormal vs normal circumstances is relevant to Allen's question about the usefulness of "normic" reasoning, Allen, 2020, this issue -- somewhere around fn 56.

[40] Gilbert, Tenney, Holland and Spellman, 2015; Petrocelli, Percy, Sherman and Tormala, 2011; Spellman, Kincannon and Stose, 2005.

Considering more counterfactuals can lead to lower causal attributions to the questioned activity; a client's defense that creates more counterfactuals to a prosecution's (or plaintiff's) theory of the case can result in smaller chance of finding causation[41]. These findings are related to work by Pennington and Hastie[42] on the story model, showing that mock jurors are more likely to find a suspect not guilty when the defense has constructed an alternative story to the prosecutor's story rather than just trying to poke holes in that story. Lawyers have plenty of anecdotal stories to this effect. Ironically, all of these findings are consistent with… relative plausibility theory.

5.   HYPOTHESIZING ABOUT HOW OTHER PEOPLE WILL THINK

Hypotheticals also permeate the legal system in the form of metacognition: thinking about what other people might think. This is a theme of Saks and Spellman's book[43]: *The Psychological Foundations of Evidence Law*. We argue that evidence rulemakers—legislators and judges—are trying to do several different things when formulating or interpreting rules. For one thing, they are, as Allen notes, juggling values[44]. Trials are not just about getting to the correct answer; they also are supposed to embody other values like fairness, finality, and efficiency. But putting those other values aside for the moment, in crafting rules of evidence law (e.g., regarding character, or hearsay, or balancing probativeness against prejudice), rulemakers are acting like applied psychologists. They may consider what effect a type of evidence «should» have on jurors, consider what effect they believe it would «actually» have on jurors, and try to figure out how the evidence should (or should not) be presented so that jurors will «get it right». But what makes us think that the rulemakers themselves will get all those steps right?[45]

Yet this is how it works. Hypotheticals and counterfactuals are important tools for understanding legal systems, mostly because legal systems are a product of human minds for engaging with human minds. It is indeed a complex adaptive system—involving people with different roles and values, incomplete and uncertain evidence that leads to known mistakes, and multiple institutions with different powers to revise the processes. At every step, we want to guess what others will think, and we may base our guesses on our intuitions, on their responses to thin hypotheticals, or on a deeper understanding of people's reasoning processes across people and contexts.

---

[41] SPELLMAN and KINCANNON, 2001; TENNEY, CLEARY and SPELLMAN, 2009.
[42] See HASTIE, 1999: 232, for a review of this research.
[43] SAKS and SPELLMAN, 2016.
[44] ALLEN, 2020 (this issue)—currently at top of manuscript p. 272—he doesn't use the word juggle but notes that «...trials pursue many objectives in addition to truth»—cite however useful
[45] Sometimes rulemakers do get it right. Read the book.

One hypothetical, weird or normal, thin or thick, tells us close to nothing about human reasoning—the true object of inquiry; but when put together wisely, they go far.

6.   BIBLIOGRAPHY

ALLEN, R.J. (this issue): «Naturalized epistemology and the law of evidence revisited»

ALLEN, R.J. and LEITER, B., 2001: «Naturalized epistemology and the law of evidence», in *Virginia Law Review*, 87(8): 1491-1550.

ALLEN, R.J. and PARDO, M.S., 2019: «Relative plausibility and its critics», in *International Journal of Evidence & Proof*, 23(102): 5-58.

CHENG, P.W. & HOLYOAK, K.J., 1989: «On the natural selection of reasoning theories», in *Cognition, 33(3): 285-313.

CHENG, P.W. and NOVICK, L.R., 1990: «A probabilistic contrast model of causal induction», in *Journal of Personality and Social Psychology*, 58(4): 545-567.

EVANS, J.S. and STANOVICH, K.E., 2013: «Dual-process theories of higher cognition: Advancing the debate», in *Perspectives on Psychological Science*, 8(3): 223-241.

FOOT, P., 1978: «The problem of abortion and the doctrine of the double effect», in *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford: Oxford University Press: 19-32.

GIGERENZER, G. and HOFFRAGE, U., 1995: «How to improve Bayesian reasoning without instruction: Frequency formats», in *Psychological Review*: 102(4), 684-704.

GILBERT, E, TENNEY, E.R., HOLLAND, C. and SPELLMAN, B.A., 2015: «Counterfactuals, control, and causation: Why knowledgeable people get blamed more», in *Personality and Social Psychology Bulletin,* 41(5): 643-658.

HASTIE, R., 1999: «The role of stories in civil jury judgments», in *University of Michigan Journal of Law Reform*, 32(2): 227-240.

HENRICH, J., HEINE, S.J. and NORENZAYAN, A., 2010: «The weirdest people in the world?», in *Behavioral and Brain Sciences*, 33(2-3): 61-83.

KAHNEMAN, D., 2011: *Thinking, fast and slow*. New York: Macmillan.

KAHNEMAN, D. and FREDERICK, S., 2002: «Representativeness revisited: Attribute substitution in intuitive judgment», in GILOVICH, T,, GRIFFIN, D. & KAHNEMAN D. (eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge: Cambridge University Press: 49-81.

KAHNEMAN, D. and KLEIN, G., 2009: «Conditions for intuitive expertise: a failure to disagree», in *American Psychologist*, 64(6): 515-526.

KAHNEMAN, D. and TVERSKY, A., 1982: «On the study of statistical intuitions», in *Cognition*, 11(2): 123-141.

MANKTELOW, K.I. and EVANS, J. ST. B., 1979: «Facilitation of reasoning by realism: Effect or non-effect?», in *British Journal of Psychology*, 70 (4): 477-488.

NIEDERMEYER, K.E., KERR, N.L. and MESSÉ, L.A., 1999: «Jurors' use of naked statistical evidence: Exploring bases and implications of the Wells effect», in *Journal of Personality and Social Psychology*, 76 (4): 533-542.

PETERSON, C.R. and BEACH, L.R., 1967: «Man as an intuitive statistician», in *Psychological Bulletin*, 68 (1): 29-46.

PETROCELLI, J.V., PERCY, E.J., SHERMAN, S.J. and TORMALA, Z.L., 2011: «Counterfactual potency», in *Journal of Personality and Social Psychology*, 100 (1): 30-46.

QUINE, W.V.O., 1951: «Two dogmas of empiricism», in *Philosophical Review*, 60 (1): 20-43.

ROEDIGER, H.L. and McDERMOTT, K.B., 1995: «Creating false memories: Remembering words not presented in lists», in *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21 (4): 803-814.

ROESE, N.J., 1997: «Counterfactual thinking», in *Psychological Bulletin*, 121 (1): 133-148.

RYLE, G., 1949: *The Concept of Mind*. New York: Harper & Row.

Saks, M.J. and Spellman, B.A., 2016: *The Psychological Foundations of Evidence Law*. New York: NYU Press.

Schauer, F., 2008: «A critical guide to vehicles in the park», in *New York University Law Review*, 83 (4): 1109-1134.

Schein, C., 2020: «The importance of context in moral judgments», in *Perspectives on Psychological Science*, 15 (2): 207-215.

Smith, E.R. and DeCoster, J., 2000: «Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems», in *Personality and Social Psychology Review*, 4 (2): 108-131.

Spellman, B.A. and Kincannon, A., 2000: «The relation between counterfactual ("but for") and causal reasoning: Experimental findings and implications for jurors' decisions», in *Law and Contemporary Problems: Causation in Law and Science,* 64 (4): 241-264.

Spellman, B.A., Kincannon, A. and Stose, S., 2005: «The relation between counterfactual and causal reasoning», in Mandel, D.R., Hilton, D.J., & Catellani, P. (eds.), *The Psychology of Counterfactual Thinking*, London: Routledge Research: 28-43.

Spellman, B.A. and Mandel, D.R., 1999: «When possibility informs reality: Counterfactual thinking as a cue to causality», in *Current Directions in Psychological Science,* 8 (4): 120-123.

Spellman, B.A. and Schnall, S., 2009: «Embodied rationality», in *Queen's Law Journal,* 35 (1): 117-164.

Tenney, E.R., Cleary, H.M.D. and Spellman, B.A., 2009: «Unpacking the doubt in "beyond a reasonable doubt": Plausible alternative stories increase not guilty verdicts», in *Basic and Applied Social Psychology,* 31 (1): 1-8, [Revised for legal application and published with response and replies as: «"This other dude did it!" A test of the alternative explanation defense», in *The Jury Expert,* 21 (4): 37-42.]

Thaler, R.H. and Sunstein, C.R., 2009: *Nudge: Improving decisions about health, wealth, and happiness*. New York: Penguin.

Thomson, J.J., 1985: «The trolley problem», in *Yale Law Journal*, 94(6): 1395-1415.

Tversky, A. and Kahneman, D., 1974: «Judgment under uncertainty: Heuristics and biases», in *Science*, 185 (4157): 1124-1131.

Wason, P.C., 1968: «Reasoning about a rule», in *Quarterly Journal of Experimental Psychology*, 20 (3): 273-281.

Wells, G.L., 1992: «Naked statistical evidence of liability: Is subjective probability enough?», in *Journal of Personality and Social Psychology*, 62 (5): 739-752.

Wright, E.F., Maceachern, L., Stoffer, E. and Macdonald, N., 1996: «Factors affecting the use of naked statistical evidence of liability», in *Journal of Social Psychology*, 136 (6): 677-688.