Chapman University

Chapman University Digital Commons

5-30-2019

Commentary: Reflections on Decision Research and Its Empiricism: Four Comments Inspired by Harrison

Nathaniel T. Wilcox
Chapman University, nwilcox@chapman.edu

Recommended Citation

Commentary: Reflections on Decision Research and Its Empiricism: Four Comments Inspired by Harrison

Comments

This book chapter was published in Michiru Nagatsu and Attilia Ruzzene, (Eds.), *Contemporary Philosophy and Social Science: An Interdisciplinary Dialogue*. Dr. Wilcox's chapter begins on page 139.

Copyright

Commentary: Reflections on Decision Research and Its Empiricism: Four Comments Inspired by Harrison

Nathaniel T. Wilcox

Generally I find Harrison's chapter cogent, interesting, and well-informed in details and particulars, and so do not speak of them. Instead, I reflect on four larger matters Harrison brings to my mind. These four matters are presented below as four separate sections, to be read as four separate and short comments (though the four sections do share a few threads).

1. Intuitions of Theorists

"In some cases simple, 'agnostic' statistical modeling is appropriate, since the experiment 'does the work of theory' for the analyst, by controlling for treatments and potential confounds," says Harrison in his introduction. In their manifesto of Bayesian statistics, Edwards, Lindman, and Savage (1963) put it in this famously entertaining way:

> It has been called the interocular traumatic test; you know what the data mean when the conclusion hits you between the eyes. The interocular traumatic test is simple, commands general agreement, and is often applicable; well-conducted experiments often come out that way. (217)

Later Edwards, Lindman, and Savage add that, "The rule was somewhat overstated by a physicist who said, 'As long as it takes statistics to find out, I prefer to investigate something else (240);'" and Ernest Rutherford allegedly said, "If your experiment needs statistics, you ought to do a better experiment." These are also the sentiments of many (perhaps most) decision researchers. I find these sentiments deeply interesting and in some respects puzzling.

What is the source of those sentiments? Consider these two quotes.

> I am about to build up a highly idealized theory of the behavior of a "rational" person ... when certain maxims are presented ... you must ask yourself ... how you would react if you noticed yourself violating them. (Savage 1954/1972; 7)
>
>   The following was offered by L. J. Savage as a criticism ... Suppose that a boy must select between having a pony *x* and a bicycle *y* and that he wavers indecisively between them [the thought experiment is further developed to a telling outcome] ... If this can happen—and the introspections of several people suggest that it can—then the strong binary model is too strong to describe these preferences. (Luce and Suppes 1965)

Theorists may sometimes use their own intuitions (or introspections) as inspiration for those "maxims" Savage alludes to above (today we generally call such maxims "axioms"). But I am not a theorist and so hesitate to speak on that matter of private inspiration: Instead, I am interested here in the persuasive role played by shared intuitions. In their writings, decision theorists reveal that intuitions play two strong roles among the theorists. Axioms are the foundation of any formal decision theory, and particularly in the case of a normative theory, a theorist frequently appeals to another theorist's intuition, as Savage does in the first quote above. A theorist will frequently state axioms in two ways: Once mathematically (for formal proofs) and once verbally to aid and persuade other theorists' intuitions (concerning the normative status, and/or the likely descriptive validity, of a proposed axiom). Second, when a theorist suggests an outcome of her thought experiment and that suggested outcome is widely endorsed by other theorists' intuitions, that *consensus of intuitions* becomes convincing evidence concerning some theory, as Luce and Suppes admit in the second quote above. Most frequently, such evidence from thought experiments is negative, suggesting a counter-example that casts doubt on the descriptive adequacy of some theory. The thought experiments are generally presented simply and transparently—largely in verbal form, perhaps with a table or two illustrating concrete sets of alternatives, but almost never more formally than that (Debreu 1960 being a notable exception). The purpose is to aid and persuade the reader's intuition that the thought experiment indeed leads to outcomes contradicting some theory. In both these cases, a consensus of intuitions or *intuition consensus* is a highly prized coin of the decision-theoretic realm.

   If you are very used to persuasion by means of transparent and simple verbal descriptions or thought experiments, you may (perhaps unfairly) devalue other less transparent kinds of evidence. Many new inferential techniques and demonstration methods—widely accepted among either classical or Bayesian statisticians—draw on our newfound bounty of computational power. The inferential techniques include simulated maximum likelihood, simulation-based Bayesian estimation, and so-called "bootstrapping" of the sampling variability of estimates. The primary demonstration method is Monte Carlo simulation, a well-established framework for examining estimators' behavior in finite samples. Here I share what one theorist said about the latter method as used in one of my own papers (Wilcox 2017):

> I am quite confident that [other scholars I respect] would not be satisfied with simulation results for a claim that could perhaps be proved analytically. I am aware of the fact that simulations are being extensively used, but I tend to believe that people resort to these methods when there is no hope of obtaining an analytical result.

Analytical results on the finite sample behavior of most nonlinear estimators are only rarely forthcoming, and only in very simple circumstances.

For whatever reasons, simulations (computation-based existence demonstrations) are a kind of evidence contemporary theorists do not find compelling. I suspect this is because simulations don't lend themselves to intuition consensus in the same way the normative status of an axiom does, or the negative outcome of a thought experiment does (nor is it an analytical proof—which, of course, the theorists find convincing too). Presented with flair, a reader can usually grasp the *results* of a simulation with no serious problem. However, if the inferential or demonstration *process* of simulation is not part of your own methodological toolbox, the process remains a kind of black box to you.

I fear that fairly or not, there is only one inferential technique that stands a chance of generating intuition consensus among the theorists: It is the interocular trauma test, which, by definition, is intuitive—"the conclusion hits you between the eyes." Edwards, Lindman, and Savage (1963) warn that "the enthusiast's interocular trauma may be the skeptic's random error. A little arithmetic to verify the extent of the trauma can yield great peace of mind for little cost (217)." Simulation-based inferences and demonstrations take us well beyond "a little arithmetic." One might say they are just hundreds of millions of instances of "a little arithmetic" assembled with care, but truly such quantity has a nontransparent quality all its own.

Aside from simulation itself, Harrison's preferred style of statistics (and it is mine too) also depends on millions of computations assembled with care: Complex, highly nonlinear likelihood functions don't get maximized without the considerable computational muscle of a computer. The output will never have the same transparency as a comparison of sample means, or the inspection of other simple sample moments. Like Harrison (see in particular his Section 4.5), I have argued that simple sample moments can be highly misleading to decision researchers (Wilcox 2008: 224–31; Wilcox 2017), but the (apparent) transparency of (potentially misleading) simple sample moments seems irresistible.

2. Estranged Siblings

Among the decision theorists, McFadden (1974, 1981) arguably had the single largest impact on *empirical* social and behavioral scientists who work with *naturally occurring* "field" data (as opposed to laboratory data): Such field researchers are the overwhelming majority of empirical economists. In his Nobel Prize lecture, McFadden (2001: 351) prominently recognized nine scholars, very much a mix of decision theorists and econometricians:

Nine other individuals who played a major role in channeling microeconometrics and choice theory toward their modern forms, and had a particularly important influence on my own work, are Zvi Griliches, L. L. Thurstone, Jacob Marschak, Duncan Luce, Amos Tversky, Danny Kahneman, Moshe Ben-Akiva, Charles Manski, and Kenneth Train.

During those remarkable twenty years from Savage ([1954] 1972) to McFadden (1974), a brisk traffic of ideas traveled between the decision theorists and the econometricians: For instance, Luce (1959) and Luce and Suppes (1965) are absolutely central to McFadden (1974), who additionally cites Block and Marschak (1960) and Tversky and Russo (1969).

In the decades since McFadden (1974), prominent decision theorists such as Fishburn (1978) and Machina (1985) also turned their talents to the apparently probabilistic nature of discrete binary choice—with no discernable impact on econometricians. And in the decades since Manski (1975), econometricians such as Cosslett (1983) and Horowitz (1992) turned their own talents to the same subject—with no discernable impact on decision research. Focus now on two papers published a quarter century ago: This gives us plenty of time to see their impact. Busemeyer and Townsend (1993) is a landmark contribution to probabilistic decision theory: It offers a very precise decision-theoretic model of both the econometric link function and index function. It is clearly influential with 569 total SSCI (Social Science Citation Index) citations, but gets zero citations from theoretical econometricians (though a handful of citations from applied econometricians). Published in the same year as Busemeyer and Townsend, most would call Klein and Spady (1993) an important milestone in econometric theory: It can free the researcher of assumptions concerning link functions—at the cost of strong assumptions (but short of linearity, and this was a major contribution to semiparametric estimation) concerning index functions. It too is clearly influential with 226 total SSCI citations, but just one (Donkers, Melenberg, and Van Soest 2001) is a decision research paper and none are decision theory papers. The sad truth is that over the quarter century since 1993, these two communities of scholars (the decision researchers and the econometricians) share about as much as the Dance and Physics Departments. Those remarkably cross-fertile years from 1954 to 1974 are well over: Econometrics and decision research went their very separate ways.

To see the ways they went, consider the probabilistic model Harrison specifies for Expected Utility Theory (EUT) in eq. (4), $\Pr(R) = f(\nabla EU)$. This is just a specific instance of the more general model $\Pr(R) = F(D(R,S \mid \theta))$, where F is any link function and $D(R,S \mid \theta)$ is any theoretical representation of the comparison between lotteries R and S—to the econometrician, the index function with parameters θ. It's fair to say many econometricians are perfectly happy to require linearity (in the parameters θ) of $D(R,S \mid \theta)$: They just want to estimate θ without assumptions concerning F, the link function. From the viewpoint of decision theory, making $D(R,S \mid \theta)$ a linear function of θ essentially takes the Prince of Denmark out of *Hamlet*: A representation without nonlinear entities in $D(R,S \mid \theta)$ just isn't worth discussing or thinking about. It's fair to say the decision researchers (Harrison and I, and at least some theorists such as Busemeyer and Townsend 1993) are fine with specific assumptions about F, if it buys

us the ability to estimate the nonlinear entities in $D(R, S \mid \theta)$ with few extra assumptions. So decision researchers and econometricians have found themselves at cross-purposes since the days of McFadden and Manski.

3. All the Horses Are Dead, Long Live the Horse Race

A decision theory τ is (usually) an axiom set such as $A^\tau = \{A_1^\tau, A_2^\tau, \ldots, A_k^\tau\}$ generating a representation (such as EUT) that applies to some prespecified set Ω of lottery pairs $\{R, S\}$. There are two empirical strategies as regards skepticism concerning such theories. The more common strategy takes a narrow focus on one or another of the axioms in A^τ, over some subset $E \subset \Omega$: Here E is a special slice of Ω, for instance, a "common ratio group" of lottery pairs in some experiment designed to interrogate A_i^τ (e.g., the Independence Axiom of EUT). Generally Ω is an infinite set, so that special slice E is a very small fraction of Ω. When this empirical strategy rejects axiom A_i^τ on subset E, we do learn something important and especially useful to decision theorists in the here and now: They may now craft a replacement for A_i^τ, hopefully leading to an improved theory.

But we need to keep clear that we learned little about the theory's performance on the set $\Omega - E$. We might be better off with a different sort of experiment: some kind of broad sampling of Ω instead of a specially contrived slice of Ω, and then a contest between the theories themselves rather than specific axioms. This less common empirical strategy (practiced by Harrison, myself and others such as Hey and Orme 1994) interrogates the *collective wisdom* of whole axiom sets A^τ by means of horse races between their representations—generally speaking with a rather less special slice of Ω as the experimental pairs. My firm conviction is that every descriptive decision theory is a dead horse walking—if we insist on its slaughter should it fail to describe every preference over all pairs in Ω for every decision maker. It is much more reasonable to race the horses (the theories, in competition with one another) and ask which ones win a noticeable fraction of the races (in other words, best explain the behavior of a noticeable fraction of our subjects on broad collections of decision problems). Harrison has this in mind when he discusses mixture models.

From both economic and evolutionary game theory we have good reason to expect living populations are *mixtures of types* (today this is well known to the point of banality). This is a primary reason (among others) why the "hypothetico-deductive" science model has limited usefulness for the *empiricism* of the biological and social sciences. I congratulate physicists for their clever selection of (mostly) the easiest possible populations (homogeneous ones) to work with, but someone has to meet the theoretical and empirical challenges of mixed populations with deep and pervasive heterogeneity. Given those types of populations, simple hypothesis-testing is potentially counterproductive. Suppose theory τ's axiom A_j^τ survives a narrow hypothesis test in set E for (say) 70 percent of subjects. But also suppose that in a competitive tests against (say) two other theories, on a broad sample of set Ω, theory τ best accounts for the behavior of just 20 percent of the subjects. I have little hesitation saying that the former test of axiom A_j^τ is at best a distraction and at worst highly misleading. When

we concentrate on relative success rather than absolute null hypothesis-testing, we're in a different world of measures of predictive success such as likelihoods, information criteria, estimated type shares in the population, and so forth. It is not a world of simple sample moments and interocular traumas.

4. Do as Theorists Say (Not as Empiricists Do)

"It cannot be said … that a rational man must behave according to the Bernoulli principle," Allais (1953: 505) concluded in the English Summary of his celebrated *Econometrica* article (in French). As Ellsberg (1961: 646) could have put the argument on Allais' behalf, "One could emphasize here … that the postulates … failed to predict reflective choices." Yet listen in today among the conferees at any decision research conference: Most regard "the Bernoulli principle" as definitive of rational decision-making under uncertainty and, in this and many other ways, we are all Savage's ([1954] 1972) children. To us, "Allais' Paradox" is a finding that subjects' decision behavior violates Subjective Expected Utility Theory (SEUT) and is therefore *not* rational (*pace* Allais and Ellsberg). As Harrison mentions, SEUT has other prominent normative Discontents (Loomes and Sugden 1982; Machina 1982; Schmeidler 1989; Epstein 1992)—Allais and Ellsberg are just the two most familiar names. But if pressed I think a majority of the conferees would agree that SEUT *is* rational choice under uncertainty.

Savage has other children: Bayesian statisticians (Box and Tiao 1973; Gelman et al. 2004), Bayesian econometricians (Zellner 1971; Geweke 2005), and Bayesian psychometricians (Kruschke 2011; Lee and Wagenmakers 2014). Edwards, Lindman, and Savage (1963) published perhaps the first manifesto of Bayesian statistics, contrasting "such procedures as a Bayesian would employ in an article submitted to the Journal of Experimental Psychology, say, and those [classical procedures] now typically found in that journal (195);" and in concluding said, "Bayesian procedures are not merely another tool for the working scientist … as we saw, evidence that leads to classical rejection of the null hypothesis will often leave a Bayesian more confident of that same null hypothesis than he was to start with (240)." They squarely address the *statistical practices of researchers* and offer the new Bayesian alternative *to those researchers' classical data analysis.* They are not talking about modeling subject behavior. Yet most citations of the manifesto borrow its mathematical results as *descriptive models of subject behavior*—to be followed, with high likelihood, by classical hypothesis-testing using the experiment's data. Uncharitable people might say normative hypocrisy has been perfected in decision research: Scholars born of Savage's seismic advance ask why subjects don't do as theorists say (not as empiricists do). Just a half dozen years ago Matthews (2011: 843) could fairly say "Judgment and decision making research overwhelmingly uses null hypothesis significance testing as the basis for statistical inference" and ask "What might judgment and decision making research be like if we took a Bayesian approach to hypothesis testing?"

Harrison argues (I think fairly) that decision research does its classical statistics with sometimes questionable rigor. But why would decision researchers do classical statistics at all, if we really believe that obedience to SEUT and Bayes' Rule is rationality

in the face of uncertainty? This question is wholly unoriginal: From conversations, I know it nags many other classical empirical economists. But in decision research, perhaps this question ought to elicit particularly sheepish grins? Or should we take our cues from Emerson and Whitman—not insisting on foolish consistency, and accepting that we contain multitudes? These are matters best addressed by philosophers, historians, and methodologists.

References

Allais, M. 1953. "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'Ecole Americaine." *Econometrica* 21: 503–46.

Block, H., and J. Marschak. 1960. "Random Orderings and Stochastic Theories of Response." In *Contributions to Probability and Statistics*, ed. I. Olkin. Stanford, CA: Stanford University Press.

Box, G. E. P., and G. C. Tiao. 1973. *Bayesian Inference in Statistical Analysis.* New York: Wiley.

Busemeyer, J. R., and J. T. Townsend. 1993. "Decision Field Theory: A Dynamic-Cognitive Approach to Decision Making in an Uncertain Environment." *Psychological Review* 100: 432–59.

Cosslett, S. R. 1983. "Distribution-free Maximum Likelihood Estimator of the Binary Choice Model." *Econometrica* 51: 765–82.

DeBreu, G. 1960. "Review of Individual Choice Behavior: A Theoretical Analysis by R. Duncan Luce." *American Economic Review* 50 (1): 186–8.

Donkers, B., B. Melenberg, and A. Van Soest. 2001. "Estimating Risk Attitudes Using Lotteries: A Large Sample Approach." *Journal of Risk and Uncertainty* 22: 165–95.

Edwards, W., H. Lindman, and L. J. Savage. 1963. "Bayesian Statistical Inference for Psychological Research." *Psychological Review* 70: 193–242.

Ellsberg, D. 1961. "Risk, Ambiguity and the Savage Axioms." *Quarterly Journal of Economics* 75: 643–69.

Epstein, L. G. 1992. "Behavior under Risk: Recent Developments in Theory and Applications." In *Advances in Economic Theory*, ed. J.-J. Laffont, 1–63. Cambridge: Cambridge University Press.

Fishburn, P. C. 1978. "A Probabilistic Expected Utility Theory of Risky Binary Choices." *International Economic Review* 19: 633–46.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. *Bayesian Data Analysis*, 2nd ed. Boca Raton, FL: Chapman and Hall.

Geweke, J. 2005. *Contemporary Bayesian Econometrics and Statistics*. Hoboken, NJ: Wiley.

Hey, J., and C. Orme. 1994. "Investigating Generalizations of Expected Utility Theory Using Experimental Data." *Econometrica* 62: 1291–326.

Horowitz, J. L. 1992. "A Smoothed Maximum Score Estimator for the Binary Response Model." *Econometrica* 60: 505–31.

Klein, R. W., and R. H. Spady. 1993. "An Efficient Semiparametric Estimator for Binary Response Models." *Econometrica* 61: 387–421.

Lee, M., and E.-J. Wagenmakers. 2014. *Bayesian Cognitive Modeling: A Practical Course.* Cambridge: Cambridge University Press.

Loomes, G., and R. Sugden. 1982. "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty." *Economic Journal* 92: 805–24.

Luce, R. D. 1959. *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

Luce, R. D., and P. Suppes. 1965. "Preference, Utility, and Subjective Probability." In *Handbook of Mathematical Psychology*, ed. R. D. Luce, R. R. Bush, and G. Eugene, 249–410. Oxford: Wiley.

Machina, M. 1982. "'Expected Utility' Analysis without the Independence Axiom." *Econometrica* 50: 277–323.

Machina, M. 1985. "Stochastic Choice Functions Generated from Deterministic Preferences." *Economic Journal* 95: 575–94.

Manski, C. 1975. "Maximum Score Estimation of the Stochastic Utility Model of Choice." *Journal of Econometrics* 3: 205–28.

Matthews, W. 2011. "What Might Judgment and Decision Making Research Be Like if We Took a Bayesian Approach to Hypothesis Testing?" *Judgment and Decision Making* 6: 843–56.

McFadden, D. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, ed. P. Zarembka, 105–142. New York: Academic Press.

McFadden, D. 1981. "Econometric Models of Probabilistic Choice." In *Structural Analysis of Discrete Data*, ed. C. Manski and D. McFadden, 198–272. Cambridge, MA: MIT Press.

McFadden, D. 2001. "Economic Choices." *American Economic Review* 91: 351–78.

Kruschke, J. K. 2011. *Doing Bayesian Data Analysis*. Burlington, MA: Academic Press.

Savage, L. J. [1954] 1972. *The Foundations of Statistics*, 2nd revised ed. New York: Dover.

Schmeidler, D. 1989. "Subjective Probability and Expected Utility without Additivity." *Econometrica* 7: 571–87.

Tversky, A., and J. Russo. 1969. "Substitutability and Similarity in Binary Choice." *Journal of Mathematical Psychology* 6: 1–12.

Wilcox, N. 2008. "Stochastic Models for Binary Discrete Choice under Risk: A Critical Primer and Econometric Comparison." In *Research in Experimental Economics Vol. 12: Risk Aversion in Experiments*, ed. J. C. Cox and G. W. Harrison, 197–292. Bingley: Emerald.

Wilcox, N. 2017. "Random Expected Utility and Certainty Equivalents: Mimicry of Probability Weighting Functions." *Journal of the Economic Science Association* 3: 161–73.

Zellner, A. 1971. *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.