



PHIL 474/673 -Natural Rationality – week 5 – February 1st

Benoit Hardy-Vallée, Department of Philosophy
University of Waterloo
<http://phi673uw.wordpress.com>

RATIONALITY AND BEHAVIORAL ECONOMICS

*“Your act was unwise,” I exclaimed “as you see
by the outcome.” He solemnly eyed me.
“When choosing the course of my action,” said he,
“I had not the outcome to guide me.*

- (Bierce, 1963, p. 68)

Rational-choice theory (RCT) formalizes individual (decision theory) and strategic rationality (game theory)

1. 3 PATHS THAT LEADS TO RCT:

(1) Mathematical enterprise: the “axioms of right reason”

The word "should" in the title of this paper has the same meaning as in the following sentences: "In building a house, why should one act on the assumption that the floor area of a room is the product and not the sum of its length and width?"; "If all A are B and all B are C, why should one avoid acting as if all C were A?" People may often act contrary to these precepts or norms but then we say that they do not act reasonably. To discuss a set of norms of reasonable behavior (or possibly two or more such sets, each set being consistent internally but possibly inconsistent with other sets) is a problem in logic, not in psychology. It is a normative, not a descriptive problem. (Marschak, 1951, pp. 493, in "Why 'should' statisticians and businessmen maximize 'moral expectation'?")

(2) Accommodation of practices, formal norms and facts (or: the Rational, the Reasonable and the Real)

- Mathematical Expectation (Arnauld & Nicole, 1964)
- Subjective Utility (Bernoulli, 1738]1954; Von Neumann & Morgenstern, 1944)
- Subjective Probability (Ramsey, 1926)
- Subjective Expected Utility (Savage, 1954)

Divergent paths:

- Alternative Expected Utility (Weighted Expected Utility, Non-Linear Expected Utility, Preference Reversals and Regret Theory; see link [1])
- Bounded Rationality (Simon)
- Ecological Rationality (Gigerenzer et al.)

(3) RCT makes explicit what is implicitly considered as normatively correct

- Norms *express* inferential rules. Syntactic norms make explicit what was semantically correct, what justifies and entails assertions (Brandom, 1994, 2000).

The numerical probabilities and desirabilities are meant to be subjective in the sense that they reflect the agent's actual belief and preferences [...]. (Jeffrey, [1965]1990, p. 1)

2 readings:

- Wittgensteinian: norms are already “in the language”
- Naturalist: norms express our naïve folkpsychology

2. INTERPRETATIONS OF RCT

(compare with interpretations of probability theory)

Descriptivism

- Describes what rational agents do or states toward which rational agents converge

Instrumentalism:

“the relevant question to ask about the “assumptions” of a theory is not whether they are descriptively “realistic,” for they never are, but whether they are sufficiently good approximations for the purpose in hand.” (Friedman, 1966, p. 15).

Normativism:

normative interpretation as a set of criteria of consistency for us to apply to our own decisions (Savage, 1954, p. 28; see also pp. 28-30, 101-103)

“Situated cognition” (Satz & Ferejohn, 1994)

RCT describes feature of agent/environment interaction, not psychological mechanisms (more details on week 8)

3. DESCRIPTIVE PROBLEMS WITH RCT

“Paradoxes” :Allais (Allais, 1953) and Ellsberg (Ellsberg, 1961)

Allais:

Choice 1: A or B?

Choice 2: A or B?

Choice 1				Choice 2			
Gamble 1A		Gamble 1B		Gamble 2A		Gamble 2B	
Winnings	Chance	Winnings	Chance	Winnings	Chance	Winnings	Chance
\$1 million	100%	\$1 million	89%	Nothing	89%	Nothing	90%
		Nothing	1%	\$1 million	11%		
		\$5 million	10%			\$5 million	10%

Thinking of probabilities as cards to pick:

Lotteries	Cards		
	1-89	90	91-100
1A	1 million	1 million	1 million
1B	1 million	0	5 million
2A	0	1 million	1 million
2B	0	0	5 million

Subjective expected utility:

1A	U(1,000,000)
1B	$[89\% \times U(1,000,000)] + [1\% \times U(0)] + [10\% \times U(1,000,000)] =$ U(1,390,000)
2A	$[11\% \times U(1,000,000)] =$ U (110 000)
2B	$[10\% \times U(5,000,000)] =$ U (500 000)

- Most people choose A in the first case and B in the second case.
- According to decision theory, If you prefer 1A to 1B, you must prefer 2A to 2B
- Why? Sure-thing principle. Since in both choices “cards” 1 to 89 are identical, they should be kept of the balance. If in choice 1 you prefer 11% chances of having 1 million (in grey) to 10% chances of having 5 million, you must prefer it in choice 2.

Experimental studies of rationality

Decision theory

- Certain bias are robustly displayed: loss-aversion, risk-aversion and ambiguity-aversion (Kahneman & Tversky, 1991, 2000; Thaler, 1980)
- Almost every (or all?) Savage axioms are not followed by subjects: the “sure-thing” principle (independence axiom) and preference transitivity, for instance, are classically violated (May, 1954; Slovic & Tversky, 1974; Tversky, 1969)
- Even subjects with utility-maximizing experiences violate rational choice theory’s prediction: traders contravene the independence axiom (List & Haigh, 2005), investors (Odean, 1998) are risk-averse and casino players have a preference reversal (Lichtenstein & Slovic, 1973). Experimented subjects, however, tend to be more in accord with decision theory.

Game theory

- Backward induction does not predict game outcome (Bosch-Domènech *et al.*, 2002; McKelvey & Palfrey, 1992)
- In games where the rational outcome is defection, such as the prisoner's dilemma (Ledyard, 1995; Sally, 1995) or the Ultimatum Game (Henrich *et al.*, 2004; Oosterbeek *et al.*, 2004; Samuelson, 2005), human subjects cooperate massively. They also prefer a null gain to an unfair gain.
- The nature of the player counts: whether subject plays the Ultimatum Game against another human or a computer, (Sanfey *et al.*, 2006), or when the players are groups (Bornstein *et al.*, 2004), their strategy tend to be closer to game-theoretical prescriptions.

4. BOUNDED RATIONALITY¹

The logic of obligation:

- Kant: “duty commands nothing but what we can do”; “ought implies can”

The encoding problem:

- Gilbert: “Yes, we should make choices by multiplying probabilities and utilities, but how can we possibly do this if we can't estimate those utilities beforehand ?” (Gilbert, 2006, p. 238)

The computational resources problem:

- Transitivity, completeness of preferences: 138 logically independent propositions (2^{138} lines) would exhaust the time resources of an imaginary computer for which each computational step would take the time for light to go across the diameter of a proton and which would be allowed to run for twenty billion years (Cherniak, 1986, pp. 93-94 & 143, n. 113).

The computational complexity problem:

- Some problems (e.g. the **traveling salesperson** (Lawler, 1985)) are not just “tough”, they are in another **computational complexity category**.
- ex: Ulysses in the Odyssey, go through 16 different cities. He had 653,837,184,000 different paths to choose from. Grötschel & Padberg's computers run for 4 days in order to find the optimal one (Grötschel & Padberg, 1993). The *function* problem is

¹ See (Aumann, 1997; Kahneman, 1994; Kahneman *et al.*, 1982; Kahneman & Tversky, 1979, 1991, 2000; Simon, 1976, 1978, 1982, 1986; Simon, [1960] 1996; Simon *et al.*, 1995; Tversky, 1969, 1975; Tversky & Kahneman, 1981; Tversky & Kahneman, 1986; Tversky *et al.*, 1990); (Shafir & LeBoeuf, 2002) and (Samuelson, 2005) are good summaries.

NP-hard, while the *decision* problem is NP-complete. Thus, heuristics are mandatory. The problem is also analogous to optimal foraging (Anderson, 1983).

- (Thagard & Verbeurgt, 1998) showed that the maximizing coherence is NP-complete, thus hardly *tractable*.

The computability problem:

- (Kramer, 1967): Turing-computability and Savage's axioms are two inconsistent sets of restrictions
- (Canning, 1992; Jones, 1982; Rabin, 1957): for some games, it is impossible to decide formally if a strategy is winning.
- See also (Binmore, 1987/88; Velupillai, 2000)

The bounds of reason

Herbert A. Simon (Nobel Prize 1978)

Complexity is deep in the nature of things, and discovering tolerable approximation procedures and heuristics that permit huge spaces to be searched very selectively lies at the heart of intelligence, whether human or artificial. (Simon, 1978, p. 13)

- procedural (not substantive) rationality
- attention and reasoning are scarce resources
- solutions must be simple and parcimonious
- the space of solutions is not pre-specified, and thus the agent must generate it
- problem-solving is sequential (thus, finite)
- perfect solutions cannot be reached, thus *satisficing* ones are acceptable
- *satisficing* solutions are the one that reach a certain aspiration level
- better solutions must be seek only when objectives are not reached
- the Ideal Rational Agent must be forgotten
- surveys, case studies, lab experiments and computer simulations must be used to study rationality.

Three senses of "bounded" (inspired from Russell & Wefald, 1991)

- by design: the designer prepares the system to be optimal under its limitations
- by adaptation: the agent adjust itself to its environment by feedback, without representing its computational resources
- by self-design: the agent takes into account its computational resources in devising behavioral policy.

Amos Tversky and Daniel Kahneman (Nobel Prize 2002)

- Emphasizes the limitation and biases of the mind.
- Analogy between decision and perception

(...) deviations of actual behavior from the normative model are too widespread to be ignored, too systematic to be dismissed as random error, and too fundamental to be accommodated by relaxing the

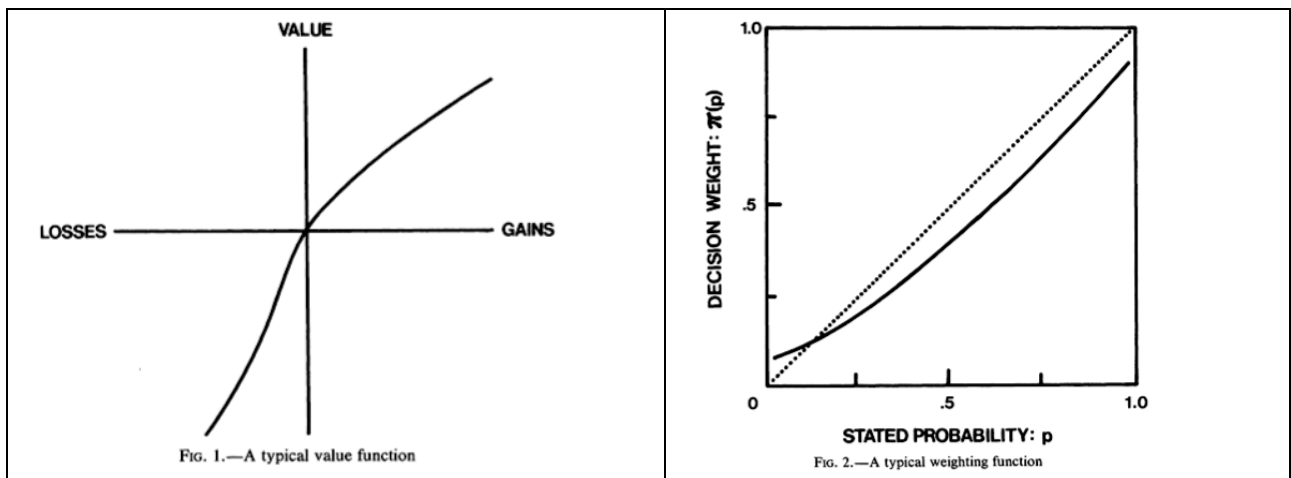
normative system. (...) the normative and the descriptive cannot be reconciled. (Tversky & Kahneman, 1986, p. s272)

Foundations of behavioral economics

- heuristics: shortcuts to solve complex problems rapidly
 - o representativeness: assumed commonality between objects of similar appearance
 - o availability: basing prediction of an outcome on the vividness of its representation rather than on actual probability
 - o anchoring: starting with a certain reference point (anchor) and then adjusting it insufficiently to reach a final conclusion.
- framing: the particular way of representing a decision context (as a gain, a loss, etc.)
- bias: cognitive bias that lead agents to deviate from rational standards
 - o ex: base-rate neglects, fundamental attribution error, confirmation bias, overconfidence, conjunction fallacy

Prospect theory: empirical theory of decision-making

- editing phase (a preliminary analysis of the offered prospects)
- evaluation phase (choosing the prospect with the highest value)
- $\pi(p)v(x)$:
 - o π : probability function
 - o p : probability
 - o v : value function
 - o x : gain



From (Tversky & Kahneman, 1986). Explains risk-aversion, endowment effect (Thaler, 1980, 1991), overconfidence.

Normative/descriptive:

«the dream of constructing a theory that is acceptable both descriptively and normatively appears unrealizable » (Tversky & Kahneman, 1986: S272)

Recent development

- Affect heuristics: non-cognitive biases
- Dual-process theory: System 1 (intuitive) vs System 2 (reasonment). System 1 triggers an intuitive judgment, and either System 2 accepts/rejects/ corrects the output of System 1 or it computes a judgement.
- Experimental evidences: (Camerer, 2000; List, 2004)
 - o List: “prospect theory adequately organizes behavior among inexperienced consumers, but consumers with intense market experience behave largely in accordance with neoclassical predictions
 - o Camerer: Prospect theory is better at making predictions and explaining data

Questions:

- Refutation of the “rationality postulate”?
- Can rationality be experimentally demonstrated/refuted?

Three reactions:

- Stich’s *pragmatism*: Humans are irrational, and empirical data should help correcting biases. (Stich, 1990, 1996)
- *Interpretivism*: (Davidson, Dennett): rationality does not amount to following decision theory. Rationality is a norm of interpretation, of “making sense”. (Davidson, 1980; Dennett, 1987). Ex: when people are risk-averse, they are not irrational, they just value certainty, and thus choose their preferred prospect. See also (Cohen, 1981)
- *Criticism* of the experiments (in (Shafir & LeBoeuf, 2002))
 - o trivialization of the findings (performance errors, insufficient monetary incentives, asking participant's justification, expertise, need for cognition)
 - o misinterpretations (conversational implications, alternative implications of terms and tasks)
 - o Inappropriate tests (computational limitations, inappropriate problem formats, inappropriate norms)

DEFINITIONS (WIKIPEDIA)

Polynomial time: the computation time of a problem where the time, $m(n)$, is no greater than a polynomial function of the problem size, n .

Decision problem: a question in some formal system with a yes-or-no answer. For example, the problem "given two numbers x and y , does x evenly divide y ?" is a decision problem.

Function problem: a function problem is a problem other than a decision problem, that is, a problem requiring a more complex answer than just YES or NO. Ex: "given two numbers x and y , what is x divided by y ?"

The traveling salesperson: Given a number of cities and the costs of traveling from any city to any other city, what is the cheapest round-trip route that visits each city exactly once and then returns to the starting city? The problem has been shown to be NP-hard, and the decision problem version ("given the costs and a number x , decide whether there is a roundtrip route cheaper than x ") is NP-complete.

P: The complexity class P is the set of decision problems that can be solved by a deterministic machine in polynomial time. This class corresponds to an intuitive idea of the problems which can be effectively solved in the worst cases.

NP: The complexity class NP is the set of decision problems that can be solved by a non-deterministic machine in polynomial time. All the problems in this class have the property that their solutions can be checked efficiently.

P = NP?: if positive solutions to a YES/NO problem can be verified quickly in polynomial time, can the answers also be computed quickly in polynomial time? (The Clay Mathematics Institute has offered a USD 1,000,000 prize for the first correct proof.)

See also <http://mathworld.wolfram.com> and (Papadimitriou, 1994)

Links

[1] Alternative Expected utility <http://cepa.newschool.edu/het/essays/uncert/alternative.htm>

References

- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine. *Econometrica*, 21(4), 503-546.
- Anderson, D. J. (1983). Optimal foraging and the traveling salesman. *Theoretical Population Biology*, 24(2), 145.
- Arnauld, A., & Nicole, P. (1964). *The art of thinking; port-royal logic*. Indianapolis,: Bobbs-Merrill.
- Aumann, R. J. (1997). Rationality and bounded rationality. *Games and Economic Behavior*, 21(1-2), 2.
- Bernoulli, D. (1738]1954). Exposition of a new theory on the measurement of risk. *Econometrica*, 22, 23-36.
- Bierce, A. (1963). *The sardonic humor of ambrose bierce*. New York,: Dover Publications.
- Binmore, K. (1987/88). Modeling rational players. *Economics and Philosophy*, 3,4, 179-214 & 179-155.
- Bornstein, G., Kugler, T., & Ziegelmeyer, A. (2004). Individual and group decisions in the centipede game : Are groups more "Rational" Players? *Journal of Experimental Social Psychology*, 40, 599-605.
- Bosch-Domènech, A., Montalvo, J. G., Nagel, R., & Satorra, A. (2002). One, two, (three), infinity, ... : Newspaper and lab beauty-contest experiments. *American Economic Review*, 92(5), 1687-1701.
- Brandom, R. (1994). *Making it explicit : Reasoning, representing, and discursive commitment*. Cambridge, Mass.: Harvard University Press.
- Brandom, R. (2000). *Articulating reasons : An introduction to inferentialism*. Cambridge, Mass. ; London: Harvard University Press.
- Camerer, C. (2000). Prospect theory in the wild. In D. Kahneman & A. Tversky (Eds.), *Choice, values, and frames* (pp. 288-300). New York: Cambridge University Press.
- Canning, D. (1992). Rationality, computability, and nash equilibrium. *Econometrica: Journal of the Econometric Society*, 60(4), 877-888.
- Cherniak, C. (1986). *Minimal rationality*. Cambridge, Mass.: MIT Press.
- Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behav Brain Sci*, 4, 317-370.
- Davidson, D. (1980). *Essays on actions and events*. Oxford, New York: Clarendon Press/Oxford University Press.
- Dennett, D. C. (1987). *The intentional stance*. Cambridge, Mass.: MIT Press.
- Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, 75, 643-669.
- Friedman, M. (1966). The methodology of positive economics. In *Essays in positive economics* (pp. 3-16). Chicago: University of Chicago Press.

- Gilbert, D. T. (2006). *Stumbling on happiness* (1st ed.). New York: A.A. Knopf.
- Grötschel, M., & Padberg, M. (1993). *Ulysses 2000: In search of optimal solutions to hard combinatorial problems* (Technical Report): New York University Stern School of Business.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., E., F., & Gintis, H. (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*: Oxford University Press.
- Jeffrey, R. C. ([1965]1990). *The logic of decision* (2nd ed.). Chicago: University of Chicago Press.
- Jones, J. P. (1982). Some undecidable determined games. *International Journal of Game Theory*, 11(2), 63.
- Kahneman, D. (1994). New challenges to the rationality assumption. *Journal of Institutional and Theoretical Economics*, 150, 18-36.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty : Heuristics and biases*. Cambridge ; New York: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263-291.
- Kahneman, D., & Tversky, A. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4), 1039-1061.
- Kahneman, D., & Tversky, A. (2000). *Choices, values, and frames*. Cambridge, UK: Cambridge University Press.
- Kramer, G. (1967). An impossibility result in the theory of decision making. In J. Bernd (Ed.), *Mathematical applications in political science, part iii* (pp. 39–51). Charlottesville: University Press of Virginia.
- Lawler, E. L. (1985). *The traveling salesman problem : A guided tour of combinatorial optimization*. Chichester [West Sussex] ; New York: Wiley.
- Ledyard, J. O. (1995). Public goods: A survey of experimental research. In J. H. Kagel & A. E. Roth (Eds.), *Handbook of experimental economics* (pp. 111-194): Princeton University Press.
- Lichtenstein, S., & Slovic, P. (1973). Response-induced reversals of preference in gambling: An extended replication in las vegas. *Journal of Experimental Psychology*, 101, 16-20.
- List, J. A. (2004). Neoclassical theory versus prospect theory: Evidence from the marketplace. *Econometrica*, 72(2), 615-625.
- List, J. A., & Haigh, M. S. (2005). A simple test of expected utility theory using professional traders. *Proc Natl Acad Sci U S A*, 102(3), 945-948.
- Marschak, J. (1951). *Why 'should' statisticians and businessmen maximize 'moral expectation'?* Paper presented at the Second Berkeley Symposium on Mathematical Statistics and Probability, Cowles Foundation Paper 53
- May, K. O. (1954). Intransitivity, utility and the aggregation of preference patterns. *Econometrica*, 22, 1–13.
- McKelvey, R. D., & Palfrey, T. R. (1992). An experimental study of the centipede game. *Econometrica*, 60(4), 803-836.
- Odean, T. (1998). Are investors reluctant to realize their losses? *The Journal of Finance*, 53(5), 1775-1798.

Oosterbeek, H., S., R., & van de Kuilen, G. (2004). Differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics*, 7, 171-188.

Papadimitriou, C. (1994). *Computational complexity*.: Addison-Wesley.

Rabin, M. (1957). Effective computability of winning strategies. In *Contributions to the theory of games iii (annals of mathematical studies 39)* (pp. 147-157).

Ramsey, F. P. (1926). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays*, (pp. 156-198). New York, London: Kegan, Paul, Trench, Trubner & Co, Harcourt, Brace and Compan.

Russell, S., & Wefald, E. (1991). *Do the right thing : Studies in limited rationality*. Cambridge, MA: MIT Press.

Sally, D. (1995). Conversations and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, 7, 58 – 92

Samuelson, L. (2005). Economic theory and experimental economics. *Journal of Economic Literature*, 43, 65-107.

Sanfey, A. G., Loewenstein, G., McClure, S. M., & Cohen, J. D. (2006). Neuroeconomics: Cross-currents in research on decision-making. *Trends Cogn Sci*, 10(3), 108-116.

Satz, D., & Ferejohn, J. (1994). Rational choice and social theory. *The Journal of Philosophy*, 91(2), 71-87.

Savage, L. J. (1954). *The foundations of statistics*. New York,: Wiley.

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology*, 53(1), 491-517.

Simon, H. A. (1976). From substantive to procedural rationality. In S. J. Latsis (Ed.), *Method and appraisal in economics* (pp. 129-148). London: Cambridge University Press.

Simon, H. A. (1978). Rationality as process and as product of thought. *American Economic Review*, 68(2), 1-16.

Simon, H. A. (1982). *Models of bounded rationality*. Cambridge, Mass.: MIT Press.

Simon, H. A. (1986). Rationality in psychology and economics. *The Journal of Business*, 59(4), S209-S224.

Simon, H. A. ([1960] 1996). *The sciences of the artificial* (3rd ed.). Cambridge, Mass.: MIT Press.

Simon, H. A., Egidi, M., & Marris, R. L. (1995). *Economics, bounded rationality and the cognitive revolution*. Aldershot, England ; Brookfield, Vt., USA: E. Elgar.

Slovic, P., & Tversky, A. (1974). Who accepts savage's axiom? *Behavioral Science*, 19, 368- 373.

Stich, S. (1990). *The fragmentation of reason : Preface to a pragmatic theory of cognitive evaluation*. Cambridge, Mass.: MIT Press.

Stich, S. (1996). *Deconstructing the mind*. New York: Oxford University Press.

Thagard, P., & Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science*, 22, 1-24.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1(1), 39-60.

- Thaler, R. H. (1991). *Quasi rational economics*. New York: Russell Sage Foundation.
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76, 31-48.
- Tversky, A. (1975). A critique of expected utility theory: Descriptive and normative considerations. *Erkenntnis*, V9(2), 163-173.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and psychology of choice. *Science*, 211, 453-458.
- Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, 59(4), S251-S278.
- Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *The American Economic Review*, 80(1), 204-217.
- Velupillai, K. (2000). *Computable economics*. New York: Oxford University Press.
- Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior* (3d ed.). Princeton: Princeton University Press.