



PHIL 474/673 -NATURAL RATIONALITY – WEEK 10 – MARCH 15

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

---

## NEUROECONOMICS AND IRRATIONALITY

---

*The discussion of the meaning of preference and the status of value may be illuminated by the well-known exchange among three baseball umpires. “I call them as I see them,” said the first. “I call them as they are,” said the second. The third disagreed, “They ain’t nothing till I call them.” Analogously, we can describe three different views regarding the nature of values. First, values exist – like body temperature – and people perceive and report them as best they can, possibly with bias (I call them as I see them). Second, people know their values and preferences directly – as they know the multiplication table (I call them as they are). Third, values or preferences are commonly constructed in the process of elicitation (they ain’t nothing till I call them). The research reviewed in this article is most compatible with the third view of preference as a constructive, context-dependent process. (Thaler & Tversky, 1990, p. 210)*

### 1 Value vs. reward

- **Reward:** refers to the immediate advantage accrued from the outcome of a decision (e.g., food, sex, or water)
- **Value:** estimate about how much reward (or punishment) will result from a decision, both now and into the future.
- *“value incorporates both immediate and long-term rewards expected from the decision. So reward is more like immediate feedback, whereas value is more like a judgment about what to expect.” (Montague et al., 2006, p. 419)*
- Maybe value is empty without reward, while reward is blind without value...

### 2 Kinds of utility

#### 2.1 Behavioral economics (Read, 2007)

- Kahneman, Wakker, and Sarin (1997) distinguish
- **Experienced** utility: hedonic content, moment-to-moment flow of pleasure or pain
  - **Instantaneous** : continuous experienced utility from sensory input, the way people feel about experiences in real-time
  - **Remembered:** memory of experienced utility, utility, which influences post-decision evaluations (e.g., regret and disappointment with a decision outcome);
  - **Predicted:** anticipation or prediction of experienced utility

- **Decision** : the utility influencing (or revealed by) the actual decision
  - *Unclear concept*: sometimes it means rational-choice theory value, sometimes objective gains and loss, sometimes the internal representation of gain vs. loss, sometimes the motivation, sometimes the value (as opposed to reward).

## 2.2 Utility in neuroeconomic: (Tom *et al.*, 2007b)

- **Predicted**: neural activity involved in the anticipation of immediate outcomes
- **Experienced**: neural representation of actual experience of gains or losses
- **Decision**: brain systems that represent potential losses versus gains when a decision is being made (the computation of ‘what to do’)
- **Multiple utility functions**: depending on the task, economic value can be represented in different areas. (‘look on your left’, ‘buy this shirt’ or ‘give her 10\$’ can be processed by different areas).

### 2.2.1 Example: loss-aversion

- vmPFC: anticipation of emotional impact (*predicted* utility) (Naqvi *et al.*, 2006)
- Amygdala: registering emotional impact (*experienced* utility) (Ibid.)
- midbrain dopaminergic neurons: computing the value of reward (*decision* utility) (Tom *et al.*, 2007a)

## 3 Neuroeconomic approach of irrationality

### 3.1 Is irrationality possible?

- Tautology problem: whatever you choose is what you prefer, thus you are always consistent
- Content of preferences problem: if an action A seems irrational, it means that we don’t understand why the agent has a preference – or a “taste” for A. *de gustibus non disputandum*
- Irrationality is possible if agents fail to want things they like. “Miswanting”

### 3.2 Bias in affective forecasting

- AF: the forecasting of one's affect (emotional state) in the future. See (Gilbert, 2006; Wilson & Gilbert, 2003). Also known as the “Affect Heuristic” (Slovic *et al.*, 2002). Predicting affective valence, intensity, duration and specific emotions.
- The construction of preferences is “a constructive, context-dependent process” (Thaler & Tversky, 1990, p. 210) - > explains *preference reversal*.
  - *an empirical regularity such that there exists a robust experimental design of lotteries for which substantial fractions of subject state prices or judgement which are opposite to the preferences expressed for or the choices made out of the respective lotteries.* (Seidl, 2002, pp. 621-622)
- **Miswanting**: cases in which people do not like or dislike an event as much as they thought they would. Wanting things that do not promote welfare, and not want things that would promote their welfare.

- **Impact bias:** overestimating the length (*durability* bias) or the *intensity* (intensity bias) of the impact of future feeling states (Gilbert *et al.*, 1998). People routinely overestimate the emotional impact of negative events ranging from professional failures and romantic breakups to electoral losses, sports defeats, and medical setbacks. They also underestimate the emotional impact of positive events such as winning a lottery.
  - (Brickman *et al.*, 1978) newly rich lottery winners rated their happiness at this stage of their life as only 4.0, (on a 6-point scale, 0-5) which does not differ significantly from the rating of the control subjects. Also surprising to many people is the fact that the paraplegics and quadriplegics rated their lives at 3.0, which is above the midpoint of the scale (2.5).
  - (Boyd *et al.*, 1990) solicited the utility of life with a colostomy from several different groups: patients who had rectal cancer and who had been treated by radiation, patients who had rectal cancer and who had been treated by a colostomy, physicians who had experience treating patients with gastrointestinal malignancies, and two groups of healthy individuals. The patients with a colostomy and the physicians rated life with a colostomy significantly higher than did the other three groups
  - Ex: The Peculiar Longevity of Things Not So Bad (Gilbert *et al.*, 2004a). The effects of extremely un/pleasant events last shorter than expected. Things that are 'not so bad' have longer effects.
  - People prefer to make changeable decisions rather than unchangeable decisions because they do not realize that they may be more satisfied with the latter. People are less susceptible to regret than they imagine, and decision makers who pay to avoid future regrets may be buying emotional insurance that they do not actually need (Gilbert & Ebert, 2002).
  - Explains loss aversion. Asymmetrical impact of losses and gains (a 100\$ loss is more 'bad' than a gain of 100\$ is 'good') is more a property of affective forecasts than a property of affective experience. (Kermer *et al.*, 2006)
- **Focalism:** Focusing too much on the occurrence in question (the *focal event*) and fail to consider the consequences of other events that are likely to occur (Wilson & Gilbert, 2003).
- **Empathy gap:** the failure to empathize or predict correctly how one will feel in the future, i.e. what kind of emotional state one will be in.
- **Immune neglect.** The failure to take into account how much one's *psychological immune system* will ameliorate reactions to negative events. People do not realize how readily they will rationalize negative outcomes once they occur.
  - Ex: Misprediction of regret. (Gilbert *et al.*, 2004b)
  - Regret: counterfactual utility/emotion/value. <sup>1</sup>

---

<sup>1</sup> Cf. (Roese & Summerville, 2005): the top six biggest regrets in life center on (in descending order) education, career, romance, parenting, the self, and leisure. (...) people's biggest regrets are a reflection of where in life they see their largest opportunities; that is, where they see tangible prospects for change, growth, and renewal.

- Problems with Regret Theory
- "What we understand by 'choiceless utility' is (...) the psychological experience of pleasure that is associated with the satisfaction of desire. **We believe that is is possible to introspect about utility, so defined**, and that it is therefore meaningful to talk about utility being experienced in choiceless situation. (Loomes & Sugden, 1982, p. 807). [Regret or rejoicing]
- Thus predicted utility based on regret minimization might be faulty since introspection is not 'direct'.
- *Much unhappiness (...) has less to do with not getting what we want, and more to do with not wanting what we like.*(Gilbert & Wilson, 2000)

## 4 Intrinsic irrationality (Berridge, 2003)

- *Irrational choice is something more than mere mistaken belief about future liking. It is choice that diverges from expectations of future liking.*

### 4.1 Kinds of utility according to Berridge:

- experienced utility (actual liking for an outcome)
- remembered utility (memory of liking in the past)
- predicted utility (expected liking for the outcome in the future)
- decision utility (manifest choice of the outcome).

### 4.2 Rationality vs Irrationality

- Rationality: Decision Utility = Predicted Utility
- Irrationality: Decision Utility > Predicted Utility (*irrational wanting*)
  - An outcome is irrationally chosen only when it is **wanted** disproportionately to its **expectation of being liked**

### 4.3 Manipulation of decision utility

- Subliminal modulation of emotions change preference ranking (Zajonc, 1998), or willingness to pay, or subjective evaluation of desirability.
- It altered the *decision* utility

### 4.4 Neural mechanisms of utility

- *Predicted* and *remembered* utility: orbitofrontal and cingulate cortex
- *Experienced* utility: nucleus accumbens, ventral pallidum and lateral hypothalamus
- *Decision* utility: subcortical mesolimbic dopamine system that projects from midbrain up to the accumbens [and prefrontal]

### 4.5 'Liking' vs. 'wanting' vs. wanting.

- ' ' = need not to be consciously represented
- **'Liking'**: hedonic impact [goal-achieving feelings]

- **‘Wanting’**: cue-triggered incentive salience. Incentive salience causes the cue and its reward to become momentarily more intensely attractive and sought. [goal-seeking cognition]
  - can be conditioned. Reinforcement learning evokes strong 'wanting' for their associated hedonic rewards and initiate goal-seeking behaviors.
- **Wanting**: explicit motivation, based on predicted utility. [goal-directed cognition]
  - “it is possible that for human minds the cognitive representation of rewards might serve to interact with dopamine systems in some circumstances” (34).
  - Goals inserted into the “reward socket” (Montague, 2006). Allow agents to value biologically arbitrary objects, achievements and actions.
- **Irrational pursuit**: the pursuit of an outcome that is not justified by *cognitive* expectations of the hedonic value of that outcome. (brain manipulation + presence of the reward)
  - “Tweaked” mesolimbic activation (injection of amphetamine that enhances dopaminergic activity) makes rats over-motivated (400%) in acquiring a conditioned reward. Nucleus accumbens dopamine specifically mediates the ability of reward cues to trigger "wanting" (incentive salience) for their associated rewards, independent of both hedonic impact and response reinforcement.(Wyvell & Berridge, 2000).
  - [Genetically modified hyperdopaminergic mice exhibit a higher ‘wanting’ without higher ‘liking’, suggesting an important role for midbrain dopaminergic system in decision utility (Pecina *et al.*, 2003).]
  - “the cue [in the mouse’s brain stimulated with amphetamine] triggered a momentary divergence of the decision utility from the predicted utility” (31). The reward cue become hyper-incentive and caused *irrational* wanting, because the motivation was not justified by predicted utility or experienced utility.
  - Drugs of abuse act as ‘brain tweaks’ that cause hyper-incentive motivated behavior. Drug addicts over-selects actions leading to drug receipt (Redish, 2004). Their brain is ‘hijacked’ by drugs.
  - **“One may ‘want’ more than one wants”**. (Berridge, 2003, p. 36)

## 5 Humean vs. AntiHumean thesis on motivation

- **Humeanism**: Desires are not beliefs. *A Humean thesis about motivation says that we are moved entirely by desire: we are disposed to do what will serve our desires according to our beliefs. If there were no desires to serve, we would never be moved more to do one thing than another. Whatever might happen then would be entirely unmotivated.* (Lewis, 1988, p. 323)
- **AntiHumean**: Desires are beliefs. Reasons are normative, or conceptual. *[Statements about what an agent prefers or desires are interpreted instead as codifying commitment to certain specific patters or practical reasoning, selected from among a wide variety of patterns that are codified by the use of other normative vocabulary.* (Brandom, 2000, p. 31)

## 6 Practical irrationality

### 6.1 Kinds of practical irrationality

- Irrational pursuit (Berridge): 'wanting' what one does not like
- Similar to *akrasia* (weakness of the will)
- *Acedia*: not wanting what one like.

### 6.2 Weakness of the will (Akrasia)

- The disposition to act contrary to one's own considered judgment about what it is best to do: the "strongest" desire does not win (ex: I should work, but there is that TV show....)
- *Year after year I read in [my journals] that I am drinking too much . . . I waste more days, I suffer deep pangs of guilt, I wake up at three in the morning with the feelings of a temperance worker. Drink, its implements, environments, and effects all seem disgusting. And yet each noon I reach for the whiskey bottle.* (Cheever, 1990, p. 54).
- *If I ask a room full of people to imagine that they've won a contest and can choose between a certified check for \$100 that they can cash immediately and a postdated certified check for \$200 that they can't cash for three years, more than half the people usually say that they would rather have the \$100 now. If I then ask what about \$100 in six years versus \$200 in nine years, virtually everyone picks the \$200. But this is the same choice seen at six years greater distance* (Ainslie, 2001, p. 33).

### 6.3 Davidson on akrasia

- (a) the agent does x intentionally
- (b) the agent believes there is an alternative action y open to them
- (c) the agent judges that, all things considered, it would be better to do y than to do x.
- Thus decision utility > predicted utility and experienced utility.

## 7 References

- Ainslie, G. (2001). *Breakdown of will*. Cambridge ; New York: Cambridge University Press.
- Berridge, K. C. (2003). Irrational pursuits: Hyper-incentives from a visceral brain. In I. Brocas & J. Carrillo (Eds.), *The psychology of economic decisions* (pp. 17-40). Oxford: Oxford University Press.
- Boyd, N. F., Sutherland, H. J., Heasman, K. Z., Tritchler, D. L., & Cummings, B. J. (1990). Whose utilities for decision analysis? *Med Decis Making*, 10(1), 58-67.
- Brandom, R. (2000). *Articulating reasons : An introduction to inferentialism*. Cambridge, Mass. ; London: Harvard University Press.
- Brickman, P., Coates, D., & Janoff-Bulman, R. (1978). Lottery winners and accident victims: Is happiness relative? *J Pers Soc Psychol*, 36(8), 917-927.
- Cheever, J. (1990). Journals. *The New Yorker*, August 13.
- Gilbert, D. T. (2006). *Stumbling on happiness* (1st ed.). New York: A.A. Knopf.
- Gilbert, D. T., & Ebert, J. E. J. (2002). Decisions and revisions: The affective forecasting of changeable outcomes. *Journal of Personality and Social Psychology*, 82(4), 503-514.
- Gilbert, D. T., Lieberman, M. D., Morewedge, C. K., & Wilson, T. D. (2004a). The peculiar longevity of things not so bad. *Psychological Science*, 15(1), 14-19.

- Gilbert, D. T., Morewedge, C. K., Risen, J. L., & Wilson, T. D. (2004b). Looking forward to looking backward. The miscalculation of regret. *Psychological Science, 15*(5), 346-350.
- Gilbert, D. T., Pines, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *J Pers Soc Psychol, 75*(3), 617-638.
- Gilbert, D. T., & Wilson, T. D. (2000). Miswanting: Some problems in the forecasting of future affective states. *Feeling and thinking: The role of affect in social cognition, 178-197*.
- Kermer, D. A., Driver-Linn, E., Wilson, T. D., & Gilbert, D. T. (2006). Loss aversion is an affective forecasting error. *Psychological Science, 17*(8), 649-653.
- Lewis, D. K. (1988). Desire as belief. *Mind, 97*(387), 323-332.
- Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal, 92*(368), 805-824.
- Montague, P. R., King-Casas, B., & Cohen, J. D. (2006). Imaging valuation models in human choice. *Annu Rev Neurosci, 29*, 417-448.
- Naqvi, N., Shiv, B., & Bechara, A. (2006). The role of emotion in decision making: A cognitive neuroscience perspective. *Current Directions in Psychological Science, 15*(5), 260-264.
- Pecina, S., Cagniard, B., Berridge, K. C., Aldridge, J. W., & Zhuang, X. (2003). Hyperdopaminergic mutant mice have higher "Wanting" But not "Liking" For sweet rewards. *J. Neurosci., 23*(28), 9395-9402.
- Read, D. (2007). Experienced utility: Utility theory from jeremy bentham to daniel kahneman. *Thinking & Reasoning, 13*(1), 45 - 61.
- Redish, A. D. (2004). Addiction as a computational process gone awry. *Science, 306*(5703), 1944-1947.
- Roese, N. J., & Summerville, A. (2005). What we regret most... And why. *Personality and Social Psychology Bulletin, 31*(9), 1273.
- Seidl, C. (2002). Preference reversal. *Journal of Economic Surveys, 16*(5), 621-655.
- Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). Rational actors or rational fools: Implications of the affect heuristic for behavioral economics. *Journal of Socio-Economics, 31*(4), 329-342.
- Thaler, A., & Tversky, R. H. (1990). Anomalies: Preference reversals. *Journal of Economic Perspectives, 4*, 201-211.
- Tom, S., Fox, C., Trepel, C., & Poldrack, R. (2007a). The neural basis of loss aversion in decision-making under risk. *Science, 315*(5811), 515-518.
- Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007b). The neural basis of loss aversion in decision-making under risk. *Science, 315*(5811), 515-518.
- Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. *Advances in experimental social psychology, 35*, 345-411.
- Wyvell, C. L., & Berridge, K. C. (2000). Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: Enhancement of reward "Wanting" Without enhanced "Liking" Or response reinforcement. *J. Neurosci., 20*(21), 8122-8130.
- Zajonc, R. B. (1998). Emotions. *The handbook of social psychology, 1*, 591-632.