What You Get is More Than What You See – or Less -- in Models of Human Decision Making

Daniel S. Levine

Abstract—Humans deciding between probabilistic alternatives may consider all or only some attributes of the alternatives. If they have a high need for cognition they may also engage in metacognitive monitoring that relates attributes of the presented alternatives to unpresented information. Neural network models of decision making data suggest brain interactions involved in high or low metacognitive monitoring.

1. WHAT YOU GET VERSUS WHAT YOU SEE

Human beings are thought to be endowed with a need to understand as much of their environment as possible. This need has been given various names by different researchers, such as the drive to comprehend [1] and the knowledge instinct [2]. Yet the knowledge instinct competes with a countervailing drive to simplify environmental inputs and minimize cognitive effort [3].

How do these competing instincts play out for decision makers (DMs) presented with alternative options, sometimes with complex descriptions? At the effort minimization end, the decision is based on a small part of the information the DM is given. In other words, for these DMs "what you get is less than what you see." At the knowledge maximization end, the decision is based not only on all of the presented information but on unpresented inferences from that information. For those DMs "what you get is more than what you see."

Elsewhere I suggested "what you get is more than what you see" as a normative ideal for mental attitudes needed to make positive social change [3]. In a similar vein, the medieval philosopher Moses Maimonides interpreted the original sin of Adam and Eve as relying on outward presentations of knowledge instead of considering its deeper implications (see [4] for discussion). Yet there are wide variations between individuals in adherence to that ideal, based on differences in cognitive traits such as need for cognition, need for closure, numeracy, and metacognitive monitoring [5]. Also, the same individual may process some information deeply and other information superficially, based on changes in context or in emotional arousal level.

Manuscript received January 30, 2020

D. S. Levine is with the Department of Psychology, University of Texas at Arlington, Arlington, TX 76019-0528, USA

Decisions between two alternatives depend heavily on how those alternatives are encoded in memory. (This is true also if there are more than two alternatives, but for simplicity we only consider binary choices in this paper.) The encoding in turn depends on which attributes of the alternatives are given attention and which attributes are ignored. Hence, modeling these choices requires a mathematical theory that integrates decision making with memory and selective attention. In previous work [6-9] I have based neural network models on a synthesis of two such theories: Fuzzy Trace Theory (FTT) and Adaptive Resonance Theory (ART).

2. REVIEW OF FTT AND ART

Fuzzy trace theory [10-12] is a unified theory of memory and decision making. It explains decisions on the basis of the way we store the presented options in memory. FTT is a dual-process theory, adducing evidence that we store events with two separate memory traces: *verbatim* and *gist* traces. Verbatim traces store stimuli exactly as they are presented; this often means exact numerical values (e.g., of money won or lost, or lives saved or lost) and probabilities. Gist traces store what the DM interprets as the essential meaning of these stimuli.

FTT also says that as people grow from childhood to adolescence and then adulthood, they gradually switch from greater reliance on verbatim traces to greater reliance on gist traces. Gist processing makes it possible for people to deal with novel situations and make ever more sophisticated generalizations and analogies. Yet gist processing also makes people more likely to remember events that did not occur and to use decision heuristics that can lead to errors or inconsistent preferences.

Broniatowski and Reyna [5] formalized FTT in a theory that proposed weighting, which could differ between individuals, between categorical gists, ordinal gists, and interval or verbatim representations. For example, they considered the Asian Disease Problem studied by Tversky and Kahneman [13]: deciding between two public health measures to combat a disease expected to kill 600 people. In one framing of that problem, the choice is between saving 200 people for sure and saving all 600 with probability 1/3 with a 2/3 probability of saving none. The categorical gist for this problem is a choice between saving some for sure and possibly saving none, which argues for the safe

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

option of saving 200. The ordinal gist is a choice between saving fewer people with higher probability and saving more with lower probability, which does not argue for either choice. The interval also makes the two alternatives equally valued because for both the expected value is 200 lives saved.

Extracting the gist of presented information amounts to categorizing the information based on relevance to the current context. Hence FTT connects naturally with neural network theories of categorization. Adaptive resonance theory (ART) [14, 15] embeds categorization in a functional network including perception, learning and conditioning, memory, and behavioral control. Recent developments of ART have made significant contact with results from neuroscience of the cortex and thalamus [16, 17].

The typical ART network is based on two layers that encode attributes and categories. The ART based models of [6-9] calculate and compare emotional values of options. In those models, the ART layers are identified with emotional regions of the brain. The lower (attribute) layer is interpreted as either amygdala or a superficial layer of orbitofrontal cortex (OFC) and the higher (category) layer as a deeper layer of OFC. Variants of the network have successfully simulated several sets of decision data involving short-term preference decisions between probabilistically presented options. The preferences are based on gist representations of the options which selectively weight different attributes encoded at the attribute layer of the network.

3. INDIVIDUAL DIFFERENCES

The Asian Disease Problem studied in [23] is a classic example of a choice that is subject to framing effects. When the choice is framed in terms of gains, as a choice between saving 200 people for sure and saving all 600 with probability 1/3 and a 2/3 probability of saving none, the choice of most participants is risk averse. That is, they prefer the sure saving of 200. Yet when the same choice is framed in terms of losses, as a choice between 400 dying for sure and a 1/3 probability of none dying with a 2/3 probability of 600 dying, the choice of most participants is risk seeking. That is, they prefer the 1/3 probability of none dying.

The different responses to gain and loss frames that Tversky and Kahneman found were based on a between-subject manipulation: participants did not see both the gain and loss frames of the same problem. Framing effects are much reduced in within-subject designs [18, 19]. This is particularly true for participants who are high in *need for cognition*. Need for cognition (NFC) is a psychological construct, based on questionnaires, that measures the tendency or inclination to engage in effortful cognitive activities.

Hence, people with high NFC tend toward "what you get is more than what you see." Another variable with the same property is numeracy, defined as facility with numerical statements and mathematical manipulations, and ability to use mathematics in the real world [20]. Numeracy and need for cognition are independent variables that both promote metacognitive monitoring, that is, continually evaluating one's choices to see that they are as consistent and sensible as possible. An example is the tendency to "edit" one's choices on a decision problem to see that they are consistent with choices on a reframing or minor perturbation of that problem.

The neural network model of AlQaudi et al. [7] and the lattice model of Broniatowski and Reyna [5] both include a metacognition parameter, varying between individuals, that measures the tendency to incorporate information that is not explicitly presented but is closely related to what is presented. This parameter, as noted above, correlates strongly with both NFC and numeracy. A high value for this parameter not only promotes consistency between gain and loss frames, but it promotes the filling in of missing information within each of those frames

Specifically, Reyna and Brainerd [11] studied the differences in framing effects between standard versions of the Asian Disease Problem and various truncated versions of the same problem. The standard gain frame is presented as: "If Program A is adopted, 200 people will be saved; If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved. Reyna and Brainerd also presented two truncated versions of the gain frame. In one of those versions the zero complement of the risky choice was removed, so that choice read simply: "If Program B is adopted, there is a one-third probability that 600 people will be saved." In the other truncated version the nonzero complement of the risky choice was removed, so that choice read: "If Program B is adopted, there is a two-thirds probability that no people will be saved." There were analogous presentations, to different participants, of the standard loss frame, the loss frame with zero complement removed, and the loss frame with nonzero complement removed.

The results of the studies in [11] were that the magnitude of the framing effect – measured as the difference in percentage of risky choice between the loss frame and the gain frame – differed between the standard and two truncated versions. Specifically, the framing effect was considerably enhanced when the nonzero complement was removed. The framing effect became insignificant when the zero complement was removed. The authors explained these findings in terms of the categorical gists from FTT, which highlighted the possibility of nobody being saved or of nobody dying. Removing the nonzero complement made those categorical gists more salient, whereas removing the zero complement made the categorical gists less salient.

Several investigators have studied brain activation patterns for choices on the Asian Disease Problem and similar probabilistic preference problems, noting how following or violating traditional frames altered brain activity. In particular, De Martino et al. [21] conducted an fMRI study of a human monetary decision task analogous to the Asian Disease Problem. De Martino and his colleagues compared activation patterns for choices that conformed to the traditional framing effect (risk seeking for losses or risk averse for gains) with choices that violated the framing effect (risk seeking for gains or risk averse for losses). These investigators found that the anterior cingulate cortex (ACC), an area involved in conflict detection, was particularly activated by choices that violated the framing effect. The orbital and ventromedial prefrontal cortex (OFC), which are involved in control and cognitive representation of emotions, showed higher activation in individuals who were less prone to the framing effect. On the other hand, frame-consistent choices tended to activate the amygdala, an area involved in primary processing of the emotional value of stimuli or events.

The next section reviews a neural network explanation for the De Martino et al. fMRI data based on ART. It also reviews the roles of the relevant brain regions in the ART- and FTT-based model of [6]

4. AN ADAPTIVE RESONANCE EXPLANATION FOR SOME IMAGING DATA

Figure 1 shows a generic ART network. The level F_1 represents attributes and F_2 represents categories. Short-term memory is encoded at the feature level F_1 and category level F_2, and learning at interlevel synapses. The orienting system generates F_2 reset when bottom-up and top-down patterns mismatch at F_1, that is, when some function that represents match between those two patterns is less some quantity called *vigilance*.

In the decision model of [6], F_1 is interpreted as amygdala, F_2 as OFC, and reset as ACC. For a participant subject to the framing effect, gist representations of prospects involving probabilities of gains and losses tend to be simplified to the four categories of "sure gain," "risk of gain or no gain," "sure loss," or "risk of loss or no loss" [12]. These gist representations favor choices of sure gains over risky gains and choices of risky losses over sure losses, in accordance with the framing heuristic. Those four gain/loss categories could be encoded at the OFC, and the actual options (e.g., gain $400 with probability 80%) at the amygdala. Resonant feedback between the amygdala and OFC (Figure 1) causes the input to be stably classified in one of those four categories, and the resulting increased F_1 activity inhibits reset (i.e., ACC) activity. Both amygdala (F_1) and OFC (F_2) are activated by the input and the interlevel feedback, but the only F_2 nodes activated are the two corresponding to the "sure gain" and "risky gain" categories; hence, total OFC activity remains small.

By contrast, consider a network corresponding to a second participant not subject to the framing effect. For such a participant, all-or-none gist representations are assumed to be weaker, and either verbatim or more nuanced gist representations stronger, than in the first participant. (An example of a nuanced gist is "almost certain gain"; see the next section of this paper.) The network corresponding to that participant has higher vigilance and may experience mismatches between those inputs and the corresponding simple categories, based on sensitivity to probability and/or magnitude information about the potential gains or losses. That mismatch means that F_1 activity no longer inhibits reset (i.e., ACC) activity. Also, there is greater activity at F_2 (i.e., OFC) than for the first participant because the all-or-none interpretation of the input is challenged and other (existing or novel) categories of gain-loss-probability configurations at that level are considered. The enhanced F_2 activity, in turn, non-specifically inhibits F_1 (i.e., amygdalar) activity.

fMRI data are notoriously difficult to interpret in mechanistic terms, which is one reason they have not often been directly reproduced in neural networks. Yet the above ART-based explanation for a specific fMRI dataset may hint at a way to interpret relative activations of brain regions in terms of the number of competing psychological representations activated at a given level. At a higher level of abstraction, say involving competing rules, this could explain greater fMRI activity in areas like the dorsolateral prefrontal cortex with greater deliberation between rules.

5. CONTEXTUAL DIFFERENCES WITHIN INDIVIDUALS

Yet other results suggest that the how carefully options are processed varies even within individuals. The same DM could rely on more nuanced gists for one option than for another on the same preference task. This was the interpretation in the model of [6] of data by Rottenstreich and Hsee [22] on differences in probability weighting based on differences in emotionality.

Rottenstreich and Hsee [22] asked some participants if they would rather obtain $50 (affect-poor) or the kiss of their favorite movie star (affect-rich), and the majority (70%) preferred the money. But when the same participants were given a hypothetical choice between a 1% probability of obtaining the $50 and a 1% probability of obtaining the kiss, the majority (65%) preferred the kiss. These researchers asked another set of participants how much they would be willing to pay for a 99% probability of obtaining a $500 tuition rebate (affect-poor) and for a 99% probability of obtaining $500 toward a trip to European tourist destinations (affect-rich: these were American students). The median price that the participants were willing to pay for the almost-certain European trip was $28 lower than the median they were willing to pay for the almost-certain tuition rebate. In both cases, the affect-rich item showed a greater enhancement of probability differences close to 0 or 1.

The model of [6] explained the kiss/money results by positing that for any desirable item the participants potentially could possess (kiss OR money), the attributes attached to the option of a certain probability of gaining that item were (a) possible gain (which has a value of 1 or 0); (b) possible non-gain (which also has a value of 1 or 0); and (c) gain probability (which has a continuum of possible values from 0 to 1). The DM's attention was assumed to shift back and forth between the two options: probability p (for some p between 0 and 1, inclusive) of kiss and probability p of money. Yet because of differences in emotional arousal, the attentional weights of the possibility attributes (a) and (b) were assumed to be larger for the kiss than for the money, and the attentional weight of the probability attribute (c) to be larger for the money than for the kiss.

Attributes (a), (b), and (c) were represented by nodes of the level F_1 of the ART module in the network (see Figure 1). Level F_2 had five nodes representing option gists: (1) sure gain; (2) sure non-gain; (3) tossup between gain and non-gain; (4) almost impossible gain; (5) almost certain gain. The more nuanced categories (4) and (5) were assumed to be more accessible when considering the money than when considering the kiss.

To my knowledge there have been no experiments testing the activations of specific brain regions in choices between equal probabilities of more and less affectively rich items. Nor have there been studies of whether the results of [22] hold in specified patient populations. It has been conjectured that the hippocampus is involved in constructing attribute weights in real time decisions via selective memory retrieval (Akram Bakkour, personal communication, 1/18/19). If that is true, amnesic patients with hippocampal damage would not show the differential probability weighting for affectively rich versus affectively poor items seen in [22].

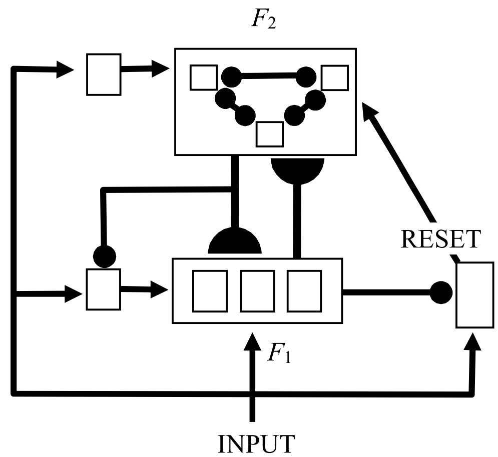

Figure 1. ART 1 architecture. Arrows denote excitation, filled circles inhibition, and semicircles learning. (Adapted from Carpenter & Grossberg, 1987, with the permission of Academic Press.)

6. CONCLUSIONS

What might distinguish neural networks in the brains of decision makers who look beyond what is directly presented to them from those who are bound only by what they directly see? The ART network analysis discussed in Section 4 of this paper argues that the "what you see is more than what you get" people show greater activation of prefrontal areas involved in cognitive and emotional control.

While the change from predominant gist to verbatim processing in adulthood is generally an advance in thinking capacity, this analysis suggests the need at time for "verbatim override" of prevailing gists if the gists are too simple. Some investigators think of the process as prevalence the fast intuitive System 1 with correction by the slow deliberative System 2 [23]. Yet the two-system description is somewhat simplistic because deliberation is not always slow and intuition is not always fast (see [24] for discussion).

The contextual difference discussed in Chapter 5 is less well understood at the neural level. If indeed some decision makers shift attention between two hypothetical alternatives and use a categorical gist encoding for one of them and an interval encoding for the other, why should that be the case? The studies that showed the phenomenon were based on students at selective universities such as Stanford, which suggests that such a shift of encoding might coexist with high need for cognition in many participants. The explanation in [6] was based on shifting levels of emotional arousal, which in the example of [22] is higher when considering the kiss than when considering the money.

Yet emotions are widely recognized to play essential roles in effective decision making [25, 26]. Hence, deliberative override of short-term arousal does not mean "the triumph of reason over emotion." Rather, it means a top-down attentional control of emotion. Such attentional control, which is both excitatory and inhibitory, is characteristic of ART networks.

References

[1] P. R. Lawrence and N. Nohria, *Driven: How Human Nature Shapes our Choices*, Jossey-Bass, San Francisco, 2002.

[2] L. I.. Perlovsky, Leonid,. "Toward physics of the mind: Concepts, emotions, consciousness, and symbols," *Physics of Life Reviews*, Vol. 3, 23-55, 2006.

[3] D. S. Levine, *Common Sense and Common Nonsense: A Conversation about Mental Attitudes, Science, and Society*. Open access e-book, first written in 1998; updated in 2018 and published by Mavs Open Press, University of Texas at Arlington Libraries. http://hdl.handle.net/10106/27541

[4] D. S. Levine and L. I. Perlovsky, Simplifying heuristics versus careful thinking: scientific analysis of millennial spiritual issues," *Zygon*, Vol 43, 797-821, 2008.

[5] D. A. Broniatowski and V. F. Reyna, "A formal model of fuzzy-trace theory: Variations on framing effects and the Allais Paradox," *Decision*, Vol. 5, 205-252, 2018.

[6] D. S. Levine, "Neural dynamics of affect, gist, probability, and choice," *Cognitive Systems Research*, Vol. 15-16, 57-72, 2012, doi:10.1016/j.cogsys.2011.07.002.

[7] B. AlQaudi, D. S. Levine, and F. L. Lewis, "Neural network model of decisions on the Asian disease problem," *Proceedings of International Joint Conference on Neural Networks 2015*, 1333-1340.

[8] D. S. Levine, K.-Y. Chen, and B. AlQaudi, "Neural network modeling of business decision making," *Proceedings of IJCNN 2017*.

[9] D. S. Levine and K.-Y. Chen, "Neural modeling of gist and verbatim in business decision making," *Proceedings of IJCNN 2018*.

[10] C. J. Brainerd and V. F. Reyna, "Gist is the grist: Fuzzy-trace theory and the new intuitionism," *Developmental Review*, Vol. 10, 3-47, 1990, doi:10.1016/0273–2297(90).90003-M.

[11] V. F. Reyna and C. J. Brainerd, "Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion. *Journal of Behavioral Decision Making*, Vol. 4, 249-262, 1991, doi:10.1002/bdm.3960040403.

[12] V. F. Reyna and C. J. Brainerd, "Numeracy, ratio bias, and denominator neglect in judgments of risk and probability," *Learning and Individual Differences*, Vol. 18, 89-107, 2008.

[13] A. Tversky and D. Kahneman, "The framing of decisions and the rationality of choice," *Science*, Vol. 211, 453-458, 1981.

[14] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Computer Vision, Graphics, and Image Processing*, Vol. 37, 54-115, 1987.

[15] G. A. Carpenter, S. Grossberg, and J. H. Reynolds, "ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network," *Neural Networks*, Vol. 4, 565-588, 1991.

[16] S. Grossberg and L. R. Pearson, L. R. (2008), "Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Toward a unified theory of how the cerebral cortex works," *Psychological Review*, Vol. 115, 677-732, 2008, doi:10.1037/a0012618.

[17] S. Grossberg and M. Versace, "Spikes, synchrony, and attentive learning by laminar thalamocortical circuits," *Brain Research*, Vol. 1218, 278-312, 2008.

[18] R. A. Le Boeuf and E. Shafir, "Deep thoughts and shallow frames: On the susceptibility to. framing effects," *Journal of Behavioral Decision Making*, Vol. 16, 77–92, 2003.. http://dx.doi.org/10.1002/bdm.433

[19] K. E. Stanovich and R. F. West, "On the relative independence of thinking biases and cognitive ability," *Journal of Personality and Social Psychology*, Vol. 94, 672–695, 2008. http://dx.doi.org/10.1037/0022-3514.94.4.672

[20]. J. M. Liberali, V. F. Reyna, S. Furlan, L.M. Stein, and S. T. Pardo, "Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment," *Journal of Behavioral Decision Making*, Vol. 25, 361-381, 2012. http://dx.doi.org/10.1002/bdm.752

[21] B. De Martino, D. Kumaran, B. Seymour, and R. J. Dolan, "Frames, biases, and rational decision-making in the human brain," *Science*, Vol. 313, 684-687, 2006.

[22] Y. Rottenstreich and C. Hsee, "Money, kisses, and electric shocks: On the affective psychology of risk," *Psychological Science*, Vol. 12, 185-190, 2001.

[23] D. Kahneman *Thinking, Fast and Slow*, Farrar, Straus, and Giroux, New York, 2011.

[24] D. S. Levine, "One or two minds? Neural network modeling of decision making by the unified self," *Neural Networks*, 120, 74-85, 2019.

[25] A. R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*, Grosset/Putnam, New York, 1994.

[26] D. S. Levine, *Healing the Reason-Emotion Split*, under contract with Routledge.