Journal of Marketing Behavior, 2015, 1: 147–156

Meta-Rationality in Cognitive Science

Keith E. Stanovich*

University of Toronto, Canada

ABSTRACT

The great rationality debate in cognitive science (Tetlock and Mellers 2002) has largely been conducted with a narrow view of human rationality in mind. A minority voice in the debate has been theorists who take a broader view of rationality — one that does not accept current desires and goals is given and that takes a longer view of decisions throughout a person's life. Schwartz's target article is clearly in the tradition of those advocating a broader view of how we conceive rationality. It has many affinities with the meta-rationality that I have previously advocated for decision science.

One way to read Schwartz's very provocative article (Schwartz 2015) is to view the thrust of his article as an argument for a broad view of rationality. To think rationally means taking the appropriate action given one's goals and beliefs, and holding beliefs that are commensurate with available evidence. But for a broad theorist, it also means adopting appropriate goals in the first place (see Stanovich 2004, Chapter 8, for a fuller discussion than I can give here). Standard instrumental rationality covers the first of these (taking the appropriate action given one's goals) and epistemic rationality covers the second (holding beliefs that are commensurate with available evidence), but when the third factor (adopting appropriate goals in the first place) is stressed, we have entered a much broader discussion of the rationality concept.

Elster (1983) deems traditional views of instrumental rationality thin theories because the individual's goals and beliefs are accepted as they are, and evaluation centers only on whether individuals are optimally satisfying desires given beliefs. Such a view represents a thin theory of rationality because

*Department of Applied Psychology and Human Development, University of Toronto, Toronto, Ontario, Canada, keith.stanovich@utoronto.ca.

Preparation of this article was supported by a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

ISSN 2326-568X; DOI 10.1561/107.00000009

it "leaves unexamined the beliefs and the desires that form the reasons for the action whose rationality we are assessing" (p. 1). The strengths of the thin theory of instrumental rationality are well–known. If the conception of rationality is restricted to a thin theory, many powerful formalisms (such as the axioms of decision theory) are available to serve as normative standards for behavior. The weaknesses of the thin theory are equally well known. In not evaluating desires, a thin theory of rationality would be forced to say that Hitler was a rational person as long as he acted in accordance with the basic axioms of choice as he went about fulfilling his grotesque desires. By failing to evaluate desires, a startlingly broad range of human behavior and cognition escapes the evaluative net of the thin theory. One could view Schwartz's target article as showing us some of the important domains that are missed.

However, developing a broad theory of rationality — one that encompasses a substantive critique of desires — does have a cost. It means taking on some very difficult issues in philosophy and cognitive science (see Stanovich 2004). It makes the evaluation of thinking more difficult and thus hinders the meliorative program of the judgment and decision making field. Modeling the psychological structure underlying broad rationality becomes a difficult task because it necessitates a deep understanding of how metarepresentation works. Let me explain why.

Dual-process theories of mind (Evans and Stanovich 2013) capture a phenomenal aspect of human decision-making that is little commented upon yet is of profound importance — that humans often feel alienated from their choices. We display what both folk psychology and philosophers term weakness of will. For example, we continue to smoke when we know that it is a harmful habit. We order a sweet after a large meal, merely an hour after pledging to ourselves that we would not. But we display alienation from our responses even in situations that do not involve weakness of will — we find ourselves recoiling from the sight of a disfigured person even after a lifetime of dedication to inclusion and equal treatment.

This feeling of alienation, although emotionally discomfiting when it occurs, is actually a reflection of a unique aspect of human cognition — the use of metarepresentational abilities to enable a cognitive critique of our beliefs and our desires. Beliefs about how well we are forming beliefs become possible because of such metarepresentation, as does the ability to evaluate one's own desires — to desire to desire differently (Dennett 1984; Dienes and Perner 1999; Jackendoff 1996; Nichols and Stich 2003; Stanovich 2004). Humans alone appear to be able to represent not only a model of the actual preference structure currently acted upon, but in addition a model of an idealized preference structure (see Kaminski *et al.* 2008; Martin and Santos 2014; Penn *et al.* 2008, 2009; Suddendorf and Corballis 2007).

Several theorists have emphasized how the metarepresentational abilities of the mind make possible a cognitive critique of our own beliefs and our

desires (e.g., Carruthers 2002; Dennett 1984; Stanovich 2004). When reasoning hypothetically, a person must be able to represent a belief as separate from the world it is representing — or entertain the idea of a goal different from the one currently being pursued. I have discussed the foundations of these metarepresentational abilities in current cognitive theory previously (Stanovich 2011). They necessarily come into play if we accept what I take to be one of the challenges in Schwartz's essay — that we should strive for what I have termed meta-rationality (Stanovich 2010).

Rational beliefs and actions are supported by strategies and knowledge that were not part of our biological endowment but instead were cultural discoveries. The development of probability theory, concepts of empiricism, mathematics, scientific inference, and logic throughout the centuries have provided humans with conceptual tools to aid in the formation and revision of belief and in their reasoning about action. As a culture, we have been engaging in a progressive cultural critique of the cognitive tools we use to act and think more rationally. This is a good thing. Principles of rational thought are not set in stone, never to be changed. In fact, the best decision making strategies will be those that are self correcting. Rationality must police its own principles if this cultural advance is to be continued. I called this the insight of meta-rationality — that all reasoning principles, even those concerned with rationality itself, must be subject to critique (Stanovich 2010).

Metarepresentational Issues in the Literature Under Discussion

Issues of meta-rationality are salient in many of the studies and research traditions that Schwartz discusses. For example, in a study by Dar-Nimrod *et al.* (2009) it was found that maximizers were more likely to expend resources in order to obtain a larger choice set. However, after choosing from the larger choice set, the maximizers who did so were less satisfied with their choice then satisficers who opted to choose from the smaller choice set. In short, the study found that maximizing is a bad strategy if one wants to maximize! Or, to put it another way, if you really want to maximize, then using raw maximizing as a heuristic is a bad choice of strategy. It can immediately be seen that the conclusion of this experiment raises issues of metarepresentation once one makes the assumption that there can be second-order knowledge of how a particular strategy fares. Someone with the knowledge of the outcome of the Dar-Nimrod *et al.* study, and who happened to be a maximizer by temperament, could use the knowledge of the study outcome to try to temper their maximizing tendencies (in the service of *actually* maximizing!).

Such a person would be, in effect, saying that "because I wish to maximize, I really should not use my automatic maximizing heuristics in these situations." The person has not really abandoned maximizing as a goal, but has simply

attained second-order knowledge that the maximizing heuristic is a bad way to achieve this goal. The situation is not unlike that of a person in a Prisoner's Dilemma game. In the iterated game, a person needs to realize that the narrowly rational, dominant response will tend to lead to a subpar outcome. Just as the person with maximizing as a goal in the Dar-Nimrod situation needs to realize that maximizing tendencies are suboptimal, the meta-rational player in a Prisoner's Dilemma needs to realize that the narrowly rational dominant response will not, in the iterated game, lead to the highest payoff. The player must realize that in this situation being narrowly rational is not broadly rational.

Such considerations of meta-rationality come into play even more in Schwartz's discussion of so-called "leaky" rationality. Schwartz gives some examples in this target article, but a previous paper by Keys and Schwartz (2007) fleshes out the issue in even more detail. For example, in a well-known sunk cost situation, the normative theory says that if a person in a hotel would turn off the movie if it were free, then the person should also turn off the movie and do something else if they had already paid $7 for it. You should not continue to do something that in the future would make you less happy because you have spent the money on it in the past. But Keys and Schwartz (2007) point out that there seems nothing wrong with feeling that the memory that you had paid the $7 might depress the enjoyment of whatever else you decided to do. You might feel bad because you had "thrown money down the drain" by not watching. Whatever the normative status of the principle of ignoring sunk costs, it seems right for people to think that "not watching the movie and regretting that you spent the $7" is a worse outcome than that of "not watching the movie". Why shouldn't people take into account the regret that they will feel if they fail to honor sunk costs?

Keys and Schwartz (2007) introduce the concept of "leakage" to help us understand this situation. Traditionally, we differentiate the decision itself from the experience of the consequences of that decision. At the time of the decision, the $7 already spent should not be a factor — so says the principle of ignoring sunk costs. But what if that $7 already spent will *in fact* affect your experience of one of the alternatives (here, specifically, the experience of the alternative of turning off the movie). If so, then the effect of the $7 (the regret at having "wasted" it) has functionally leaked into the experience of the consequences — and if it is in fact part of the consequence of that option, then why should it be ignored?

In the target article, Schwartz cites studies where (seemingly) irrelevant factors that are used to frame the choice actually leak into the experience of the alternative chosen. He cites studies in which beef labeled as "75% lean" was experienced as tasting better than beef labeled "25% fat" and in which people performed better after drinking an energy drink that cost $1.89 than after consuming the same drink when it cost only $0.89. Again, the way

the alternatives were framed leaked into the experience of the alternatives. So one argument is that framing effects in such situations are not irrational because the frame leaks into the experience of the consequence. In fact, Keys and Schwartz (2007) point out that if leakage is a fact of life, then the wise decision-maker might actually want to take it into account when making decisions.

Consider another example of regret as a leakage factor. Keys and Schwartz (2007) discuss the example of standing in the grocery store line and suspecting that the neighboring line would move faster. What should we do, stay or switch? On the one hand, our visual inspection of the neighboring line and the people in it leads us to suspect that it would move faster. Why would we ever hesitate if this were our judgment? Often, the reason we do in fact hesitate is because we can recall instances in the past where we have switched lines and then observed that our original line actually ended up moving faster. We want to kick ourselves when this happens — we regret our decision to switch. And we tend to regret it more than when we fail to switch and the neighboring line does indeed move faster. If we take this anticipatory regret into account, we might well decide to stay in the line we are in even when the neighboring line looks like it will move faster.

In the grocery line and the movie examples, anticipated regret leads us to take actions that would otherwise not be best for us (in the absence of such anticipation). One response to these choices is to defend them as rational cases of taking into account aspects of decision framing that actually do leak into the experienced utility of the action once taken. Another response is to argue that if regret is leading us away from actions that would otherwise be better for us, then perhaps what should be questioned is whether the regret we feel in various situations is appropriate.

This response — that maybe we should *not* let aspects of how the choices are framed leak into our experience — Keys and Schwartz (2007) call "leak plugging". That leak plugging may sometimes be called for is suggested by another example that they discuss — that students think that if they change their responses on a multiple-choice test that they are more likely to change a correct response into an incorrect one than vice versa. Keys and Schwartz (2007) point out that this belief is false, but that it may be a superstition that arose to help prevent regret. That there is another response to regret other than avoidance is suggested by a question that we might ask about the situation surrounding the multiple-choice superstition: Are people better off with lower grades and reduced regret, or are they better off with some regret but higher grades?

The multiple choice example thus suggests another response to decision leakage of contextual factors — that rather than simply accommodating such leakage into our utility calculations, we might consider getting rid of the leakage. In short, maybe the most rational thing to do is to condition ourselves to avoid

regret in situations where we would choose otherwise without it. Without the regret we could freely and rationally choose to turn off the movie and enjoy an activity that is more fulfilling than watching a boring film. Without the regret we could change to whichever grocery line looked more promising and not worry about our feelings if our predicted outcome did not occur. Schwartz's point that leakage is less when we are making decisions for others, is consistent with the prescription that probably there is a lot of this leakage in our own decisions that should be "plugged".

Note that a decision to condition ourselves to avoid regret in the movie example would represent a more critical use of the rational principle of avoiding the honoring of sunk costs. It would reflect a use of the sunk cost principle that was informed by a meta-rational critique — one that took a critical stance towards the rational principle rather than applying it blindly. A first-order use of the sunk cost principle would apply it no matter what and — given the natural structure of human psychology — would sometimes result in lower experienced utility because the blind use of the principle fails to account for regret. A critical stance toward the principle would recognize that sometimes it leads to lower experienced utility due to the unaccounted-for regret. But, as a further step in meta-rational analysis, the regret itself might be critiqued. The sunk cost principle comes into play again in reminding us that, without regret, turning off the movie is the better choice. If, at this point, we decide to endorse the sunk cost principle, it is in a much more reflective way than simply blindly applying it as a rule without a consideration of human psychology. The decision to alter our psychologies in light of the rule would in a sense be a second-order use of the rule, one that represented a meta-rational judgment.

This aspect of meta-rationality is in effect asking about the appropriateness of our emotional reactions to a decision. If we deem these reactions appropriate, then they must be factored in. Sometimes, however, we will deem the emotions less important than our other goals. We will want the better grades, the better line at the grocery store, and the activity that is better than the boring movie — and we will want all of these things more than we value avoiding regret. In this case, we revert to the traditional normative rule — but only after having engaged in meta-rational reflection.

Situations that are repeated are more likely to be the ones where we might want to plug leakage and target some of our emotions for reform. Schwartz discusses a person who is afraid of elevators. Someone afraid of elevators might be perfectly rational, on a narrow analysis of a *particular occasion*, in taking the stairs even though the stairs are slower because they have factored in the negative utility of their fear while riding in the elevator. However, as Schwartz notes, such a person living and working in New York City, might well think of accepting some therapy in the service of ridding themselves of this fear. What might look rational on a given single occasion, might seem very suboptimal from the standpoint of a *lifespan* filled with similar activities.

Financial decisions that cumulate have a similar logic. Suppose you are the type of person who is affected by friendly salespeople, you tend to buy products from those who are friendly. Furthermore, suppose that there is leakage from decision to experience regarding this factor — you actually enjoy products more when you have purchased them from friendly people. Clearly though, given the logic of our market-based society, you are going to end up paying much more for many of your consumer goods throughout your lifetime. Here, a lifetime and a single case tend to look very different. You pay 25 cents more for a coffee from the Bean People tomorrow because you like them better than the Java People. No problem. But you might answer differently if calculations were to show that buying from friendly people will cost you a compounded return of $175,667 in your retirement fund over a lifetime. With this information, you might decide to plug the leakage and stop responding to the "friendly factor" in your future financial decisions.

An actual consumer example comes from the "extended warranties" that are sold with many appliances. At the time of each individual purchase, these small-scale insurance contracts may give us some reassurance and comfort. But consumer magazines routinely report that, when aggregated, these are very bad products. That is, across a number of such contracts, the return to the consumer is very low — much more is spent in premiums than is actually returned by making a claim on the warranty. Of course, on one particular purchase, buying the warranty might have positive utility — not because of any money saved, but because it reduces the negative utility of the anxiety we feel at the time of purchase. Nonetheless, however comforting the warranty is in the case of *this particular* appliance, across several such appliances they are a very bad deal. Thus, the consumer is better off by trying to get rid of the purchase anxiety that leads them to buy the warranty each time.

Meta-Rationality and Norms

These examples show the more delicate interplay between normative rules, individual decisions, and a long-term view of one's goals and desires that takes place when meta-rationality rather than a thin instrumental rationality is our concern. My view of this seems to be completely congruent with Schwartz's summary statement that "On this global view of rationality, which tries to integrate over multiple decisions and experiences, the emphasis is on making people more rational, sometimes by altering substantive characteristics of the person as decision maker and experiencer, and sometimes by altering formal procedures for evaluating options and making decisions. What rationality requires will depend on both the decision maker and the context within which the decision is made. The trick may be to value formal principles of rationality, but not take them too seriously" (p. 30).

Some may read Schwartz's essay as advocating a Panglossian position on human rationality, but I do not. I refer here to my (Stanovich 1999, 2004) distinction between the Meliorists, who assume that human reasoning is not as good as it could be and that thinking could be improved, and the Panglossians who argue that an assumption of perfect human rationality is the proper default position to take. The latter posit no difference between descriptive and normative models of performance, because human performance is actually normative. The former feel that education and the provision of information could help make people more rational — could help them to further their goals more efficiently and to bring their beliefs more in line with the actual state of the world.

Engaging in a meta-rational critique can be viewed as one way of reconciling the positions of the Panglossians and the Meliorists. Meliorists stress the importance of normative rules, the necessity of following them, and correcting ourselves when we deviate from them. Panglossians stress that deviations might be only apparent and that sometimes they are the result of normative rules that are inappropriately applied. An emphasis on meta-rationality recognizes the importance of normative rules while at the same time stressing, as Schwartz does, that such rules are very much open to critique.

Meta-rationality fuses the views of the Meliorist and Panglossian by revealing the incompleteness of both views. An unreflective Meliorist is too quick to apply a blanket normative rule to a specific situation that may have alternative interpretations and subtle contextual factors that might leak into the experience of the consequences. Panglossians are of course quick to point out that there may be rational alternative interpretations of the task and that emotions at the time of decision might leak into experience. But Panglossians sometimes fail to take a broader view of life — one that would examine how certain responses may have cumulative effects over time. Panglossians often fail to see how the hostile environment of many market-based societies will exploit the "alternative interpretation" of the decision-maker (Stanovich 2012). A broader view of life, one that recognized hostile environments and that recognized the cumulative effect of repeated instances, might dictate more attention to normative rules. Meta-rationality demands a broad view of life, a concern for task construal and the role of emotions in decision, *as well as* a concern for the expected utility of the outcome. This view seems not inconsistent with the entire thrust of what Schwartz argues in this thoughtful essay.

References

Carruthers, Peter (2002), "The Cognitive Functions of Language," *Behavioral and Brain Sciences*, 25, 657–726.

Dar-Nimrod, Ilan, Catherine D. Rawn, Darrin R. Lehman, and Barry Schwartz (2009), "The Maximization Paradox: The Costs of Seeking Alternatives," *Personality and Individual Differences*, 46, 631–5.

Dennett, Daniel C. (1984), *Elbow Room: The Varieties of Free Will Worth Wanting*, Cambridge, MA: MIT Press.

Dienes, Zoltan and Josef Perner (1999), "A Theory of Implicit and Explicit Knowledge," *Behavioral and Brain Sciences*, 22, 735–808.

Elster, Jon (1983), *Sour Grapes: Studies in the Subversion of Rationality*, Cambridge, England: Cambridge University Press.

Evans, Jonathan St. B. T. and Keith E. Stanovich (2013), "Dual-process Theories of Higher Cognition: Advancing the Debate," *Perspectives on Psychological Science*, 8, 223–41.

Jackendoff, Ray (1996), "How Language Helps Us Think," *Pragmatics and Cognition*, 4, 1–34.

Kaminski, Juliane, Josep Call, and Michael Tomasello (2008), "Chimpanzees Know What Others Know, But Not What They Believe," *Cognition*, 109, 224–34.

Keys, Daniel J. and Barry Schwartz (2007), ""Leaky" Rationality: How Research on Behavioral Decision Making Challenges Normative Standards of Rationality," *Perspectives on Psychological Science*, 2, 162–80.

Martin, Alia and Laurie R. Santos (2014), "The Origins of Belief Representation: Monkeys Fail to Automatically Represent Others' Beliefs," *Cognition*, 130, 300–8.

Nichols, Shaun and Stephen P. Stich (2003), *Mindreading: An Integrated Account of Pretence, Self-awareness, and Understanding Other Minds*, Oxford: Oxford University Press.

Penn, Derek C., Keith J. Holyoak, and Daniel J. Povinelli (2008), "Darwin's Mistake: Explaining the Discontinuity Between Human and Nonhuman Minds," *Behavioral and Brain Sciences*, 31, 109–78.

Penn, Derek C., Keith J. Holyoak, and Daniel J. Povinelli (2009), "Universal Grammar and Mental Continuity: Two Modern Myths," *Behavioral and Brain Sciences*, 32, 462–4.

Schwartz, Barry (2015), "What does it Mean to be a Rational Decision Maker," *Journal of Marketing Behavior*, 1(2), 113–45.

Stanovich, Keith E. (1999), *Who is Rational? Studies of Individual Differences in Reasoning*, Mahweh, NJ: Erlbaum.

Stanovich, Keith E. (2004), *The Robot's Rebellion: Finding Meaning in the Age of Darwin*, Chicago: University of Chicago Press.

Stanovich, Keith E. (2010), *Decision Making and Rationality in the Modern World*, New York: Oxford University Press.

Stanovich, Keith E. (2011), *Rationality and the Reflective Mind*, New York: Oxford University Press.

Stanovich, Keith E. (2012), "Environments for Fast and Slow Thinking," *Trends in Cognitive Sciences*, 16(4), 198–9.

Suddendorf, Thomas and Michael C. Corballis (2007), "The Evolution of Foresight: What is Mental Time Travel and is it Unique to Humans?" *Behavioral and Brain Sciences*, 30, 299–351.

Tetlock, Philip E. and Barbara A. Mellers (2002), "The Great Rationality Debate," *Psychological Science*, 13, 94–9.