Review

Keith Stanovich

Rationality and the reflective mind (Oxford University Press, 2010)

The relationship between the concept of rationality and the so-called Dual Process Theory (hereafter, DPT) has always been discussed during the course of development of the theory. This makes good sense, since the heuristics and biases tradition is a birthplace of DPT. Why do some participants fall prey to obvious mistakes in, say, the Linda problem and others not? According to DPT, the general answer is as follows: there are two information-processing systems (Systems 1 & 2) in our mind, and we often employ either or both of them in problem-solving tasks. System 1 is activated when we make a quick-and-dirty, "intuitive" response, while System 2 is working when we make a slow but deliberative response. And some participants rely on the intuitive processes while others "think twice" and let the deliberative or perhaps "rational" mind (System 2) override it. One might follow stereotypical thinking at one time and judge that Linda is more likely to be a feminist bank teller than a bank teller. But the deliberative processes may come to the rescue and she might change her mind to conclude that her initial answer would violate the law of conjunction in probability theory.

Keith Stanovich has made significant contributions to the development of DPT in this area. One is his paper with Richard West ('Individual differences in reasoning: Implications for the rationality debate?', *Behavioral and Brain Sciences* **23**, 645–665, 2000). In the paper Stanovich observed a substantial amount of individual differences in their performance in the heuristics and biases experiments ——some do give the "rational" answers while others not—— and noted that there is correlation between the proportion of the rational answers and the level of intelligence in many cases. From this he inferred that people of high intelligence are those better at overriding the intuitive responses coming from System 1 to produce rational and deliberative responses. Stanovich thereby connected the overriding function of System 2 with intelligence, and took this correlation as evidence against the Panglossian interpretation of the results of the heuristics and biases experiments espoused by evolutionary psychologists.

Rationality and the reflective mind (Oxford University Press, 2010) is his latest contri-

bution to the Dual Process Theory and our understanding of rationality. Readers would find several developments of DPT and his conception of rationality in this book. Firstly, Stanovich gives a more precise description of the original two types of processes. Take Type-2 processes (Stanovich is reluctant to use the "system" terminology in this book because it gives an impression that DPT implies that there are only two Systems (as tokens) in our mind; he notes that DPT is about *kinds* of mental processes, not tokens (p. 19)). Stanovich divides the functions of Type-2 processes into two categories. The first function is to *override* the responses from Type-1 processes when it is appropriate to do so. This means that Type-2 processes have inhibitory mechanisms in them. But this is not enough, because inhibition does not produce any response by itself. Thus Type-2 processes need to have some mechanisms to give alternative responses. Stanovich believes that it is what he calls *cognitive decoupling* that does a significant part of that job. Cognitive decoupling is to create and maintain a secondary mental representation when one has a primary representation about the world (p. 49). The primary representation is in a direct connection to the world typically through perception, but the secondary representation is not. In other words, cognitive decoupling is something which enables us to imagine something unreal. He then points out that the capacity of working memory, which is strongly correlated with the level of fluid intelligence (thereby Type-2 processes), is actually more about maintaining one's attention in spite of distraction ——this is what decoupling helps us do by keeping the secondary representation—— than just about the volume of memory storage. It is in this sense that this type of processes is partially responsible for our rationality. Stanovich explains the working of Type-2 processes in terms of working memory and cognitive decoupling.

But Stanovich does not stop here. Unlike his earlier paper, he argues that besides intelligence, a group of *thinking dispositions* are important in thinking rationally. Those dispositions include open-minded thinking, need for cognition (the tendency to think a lot), and consideration of future consequences (p. 35). In a nutshell, they involve our capacity to initiate and regulate the original Type-2 processes. Why are those thinking dispositions important? One reason is that Type-2 processes as conceptualized above do not give a complete explanation of our performance in some of the heuristics and biases quizzes. Consider myside bias. When psychologists ask university students to evaluate an argument supporting death penalty, for example, the students overestimate its relevance

if they also support death penalty, and *vice versa*. In other words, one tends to evaluate an argument impartially according to her position on an issue. And importantly, whether one falls prey to this bias is not correlated to the level of her intelligence: even intelligent people evaluate an argument at hand differently whether it confirms or disconfirms their position. Moreover, even in the quizzes where one's performance does correlate with her cognitive ability as measured by intelligence tests, its correlation efficients are generally rather modest (for example, from 0.25 to 0.35 in various probabilistic reasoning tasks (p. 123; see also Chapter 8)), and whether to have relevant thinking dispositions is another predictor of one's performance after controlling her intelligence.

This means that having appropriate thinking dispositions and intelligence are both necessary for us to think rationally. Stanovich calls the mental processes behind those dispositions *the reflective mind* and the original Type-2 processes *the algorithmic mind*. This is because Stanovich believes that cognitive abilities measured in the intelligence tests are those one exerts when a task is defined so clearly that there is no need for interpretation and thereby one can exhibit their *maximal* cognitive performance (p. 39). But when the problem is not defined clearly enough, one needs to initiate the deliberative processes in the first place (this is a job for the reflective mind).

Based on the distinction between the algorithmic and reflective minds, Stanovich re-examines the concepts of rationality and intelligence. There have been fierce debates on intelligence in psychology and other related fields. One of them has focused on the malleability of intelligence in individuals or on the population level. But, despite the differences in views, both sides on this debate commonly assume that the concept of intelligence covers all of (or most of) our cognitive capacities (p. 122). The findings cited in this book, however, provide an alternative picture; for the algorithmic mind alone does not let us think rationally in various reasoning quizzes. Perhaps the concept of intelligence misses an important component of our rationality. If this is true, it will demand a shift in the focus when we discuss the concept of intelligence. Even if intelligence is not genetically very malleable, as some fear, this does not mean that intelligent people always act rationally. Rather, they may act irrationally and fail dramatically in real life. Intelligent people calculate faster and more accurately than otherwise *if* they can let the algorithmic mind work fully. But if their mind does not initiate it in the first place, they would fall prey to all kinds of fallacies and biases (this point is extensively discussed in

his another book *What intelligent tests miss* (Yale University Press, 2009)).

This is how Stanovich adds substantially to our understanding of rationality. Now one could ask a couple of critical questions on his attempts. One obvious question is on cost-benefit trade-offs of his decision to introduce the reflective mind, the third component of the mind, into DPT, which now Stanovich calls a *tripartite* theory of mind. Stanovich replies that this does not mean that he entirely abandons DPT, because the important point DPT makes to psychologists and philosophers is that there is more than one kind of process in our mind (p. 33). Whether there is one or many is important, while whether two or three is not. I agree with him in this respect. However, this should not be the end of the story. Now the pressing questions are whether there is any guarantee that we won't go beyond it to have the four-part theory of mind (and it goes beyond that again...), and what is the condition to call one kind of process distinctively different from other, i.e., individuation criteria for the kinds of processes. Although Stanovich discusses the way in which the reflective mind is different from the algorithmic mind, he does not discuss such criteria. This is something missing in the book, although this may ask too much for psychologists, because it may be a kind of job philosophers should be better at doing.

Another concern is with the fact that DPT is originally supposed to integrate various fields of psychology. Discussions in this book entirely focus on the heuristics and biases literature. But as Jonathan Evans notes, DPT can be seen as an integration of findings from various subfields of psychology such as social psychology. Then one may wonder if this tripartite structure can be seen in those fields. For example, can we observe similar dispositions to initiate and regulate the algorithmic mind when we form an impression on a person (Brewer, M., 'A dual process model of impression formation', *Advances in Social Cognition* **1**, 1–36, 1988)?

These quibbles aside, Stanovich's book still serves as a valuable contribution to the scholarship. His distinction between rationality and intelligence will not only have important theoretical implications to the philosophical debates on rationality, but also practical implications to our society, because we use the scores of intelligence tests (or substitutes like SAT) for screening in various contexts, as Stanovich discusses in Chapter 9. Any readers interested in rationality and intelligence would enjoy reading this book.

（AMITANI Yuichi, Faculty of Bioindustry, Tokyo University of Agriculture. Email: yuiami@gmail.com）