An acid test of human rationality:

You are shown a set of four cards placed on a table, each of which has a number on one side and a letter on the other. Visible faces of the cards feature K, 5, 8, A. You are told that the rule according to which letters and numbers are matched in each card is as follows: if a card has a vowel letter on one side, then its reverse features an even number. How many and which cards should you turn over in order to test this claim?

A naive solution would be to act by brute force, and turn over all 4 cards—but this is clearly not the most efficient one. Upon some reflection, most people choose A and 8, opting to observe an even number and a vowel, respectively on the reverse sides of these cards. This is more efficient decision, yet it is incorrect. The claim does not say that a vowel letter must be on the other side of an even numbered card—in fact, what is over an even number is unrestricted. Yet the claim implies (by contraposition) that if the number is not even (i.e. odd), then it must not correspond to a vowel. Hence, the right and efficient answer to the riddle is to turn over the two cards featuring A and 5.

A task of finding quick and efficient acid tests of some claim, strategy or plan abound in real life, including business and public projects, political or personal affairs. How good are we humans at finding such solutions? And in general, how justifiable is the claim of traditional (neoclassical) economics that people are rational? Psychologist Keith Stanovich has been one of the most active researcher of the limits of human rationality. Inter alia, he coined the notion of two cognitive Types, which have been widely popularized by Daniel Kahneman’s bestseller Thinking Fast and Slow (Kahneman 2011) as System 1 and System 2. System 1 thinks fast: it is responsible for Unconscious Reasoning, Implicit, Automatic, Low Effort, Default Process Associative, Non-Logical, Parallel processing of information. System 2 is slow: it deals with Conscious, Reasoning, Explicit, Controlled, High Effort, Inhibitory, Rule-Based, Abstract, Logical, Serial decisions. The claim that human decisions are made as a combination of these two constitutes the essence of the dual-process theory. According to it, people default to cognitive processing mechanisms of low computation expense. Whenever possible, it is System 1 that will act and make decisions based on unconscious, automated decision rule—such as ‘turn over all cards’. However, facing an unknown challenge, this simple intuition may be judged inappropriate—in this case, System 2 will take over, but still defaulting to particular reasoning routines. In the four cards task, Evans (2006) presents evidence that this is exactly what happens when subjects reach a more involved decisions of turning over two cards, including the case incorrect choice of A and 8; as well as in many other decisions, such as widely famous framing effect. This reliance to decision routines, or a single focal model that triggers all subsequent thought in a given context (Stanovich e.a., 2016, p.100) is termed serial associative cognition, and is common to both System 1 and System 2.
The above evidence is suggestive that the two Types of reasoning (or two Systems) alone are not always the best advisors and guards to human rationality. But is it sufficient to guide us the humans to make right decisions? In their 2016 book, Stanovich, West and Toplak made a step further, and offer a tripartite model of human reasoning. A crucial ingredient to that theory is the set of metacognitive capacities to decouple the routine representations by overriding them through mental simulation and hypothetical thinking: what if the intuitive routines are incorrect? Can I think of an alternative explanation? Shall I look at the task from a different perspective? Capability to do that is not synonymous to analytical skills as captured by the IQ-type tests, nor to Frederick’s Cognitive Reflection Test (Frederick 2005). Following Perkins (1995), Stanovich and co-authors label these capabilities mindware, establish their connection to the other types of reasoning, human rationality and intelligence (Baron 2005). A synthetic scheme of the tripartite model of cognition is provided on Figure 2.4. (p.107); in subsequent chapters the authors discuss the potent of its use in different tasks, especially those leading to potential biases, as well as characterize limitations and pitfalls of its application.

Besides theoretical conceptualization of what makes a good decision, the authors develop a method of testing human rationality. In The rationality quotient they present the result of several year’s work to construct a Comprehensive Assessment of Rational Thinking (CART test) which in full version included 4 domains and 20 subtests, covering processing requirements (Detection, Sustained Override, Hypothetical Thinking), knowledge of prescriptive models (Probabilistic Numeracy Subtest, Risk knowledge Subtest, etc.), avoidance of contaminated mindware (for instance, anti-scientific attitudes), and thinking dispositions that foster thorough and prudent thought, unbiased thought, and knowledge acquisition.

Trial testing and calibration of the subtests were carried out on several hundred subjects. An entire part of the book is devoted to statistical validation of CART, in which two empirical results seem to be most notable. One if that the study rejects again the hypothesis of high correlation between capacity of Rational Thinking and IQ –thus, these two concepts belong to different realms of human abilities. Perhaps more importantly, they notice high correlation between CART and the actively open-minded thinking scale, suggesting that Rational Thinking is part of a general body of human aptitudes, such as future orientation and differentiation of emotions (p.226).

Of course, this test is not a finished and ultimate solution as to how human rationality should be judged upon –but this is a significant step towards its understanding and measurement. This book opens the door to many further studies, is highly thought-provoking and suggestive of the new ways of assessing a broad range of human activities, ranging from educational programs to policy, welfare and other socio-economic aspects of human life.

**References**


