Cell PRESS

Abnormality, rationality, and sanity

Ralph Hertwig[1] and Kirsten G. Volz[2]

[1] Max Planck Institute for Human Development, Center for Adaptive Rationality, Lentzeallee 94, 14195 Berlin, Germany
[2] Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Strasse 25, 72076 Tübingen, Germany

A growing body of studies suggests that neurological and mental abnormalities foster conformity to norms of rationality that are widely endorsed in economics and psychology, whereas normality stands in the way of rationality thus defined. Here, we outline the main findings of these studies, discuss their implications for experimental design, and consider how 'sane' some benchmarks of rationality really are.

How brain damage 'cures' ambiguity aversion

A remarkable pattern has begun to emerge from recent studies in neuroscience and neuroeconomics: sometimes individuals with mental illnesses or damage to specific brain regions are more likely than the hale and hearty to adhere to classic benchmarks of rationality. Take, for instance, the classic Ellsberg paradox, an instance of choice behavior that challenges (subjective) expected utility theory. Suppose two urns each contain 100 balls. The risky urn contains 50 red and 50 black balls. The ambiguous urn contains red and black balls in unknown proportions. When individuals who stand to receive a prize for drawing a red ball are invited to choose between the two urns, they generally prefer the risky to the ambiguous urn. When they are subsequently promised a prize for drawing a black ball, their aversion to the ambiguous urn remains. This persistent preference, which leads humans to stray from the axioms of standard decision theory, has been interpreted as ambiguity aversion [1]. If individuals prefer to draw a red ball from the risky rather than the ambiguous urn, then their subjective probability of drawing a red ball from the ambiguous urn must be <0.5, in which case the complementary probability of drawing a black ball must be >0.5. Therefore, they should prefer to draw a black ball from the ambiguous rather than the risky urn, but most do not. However, the 'right' kind of brain damage seems to cure humans of ambiguity aversion. Patients with lesions to the orbitofrontal cortex (OFC), for example, have no systematic preference for the risky urn, which 'is behaviorally abnormal but is consistent, ironically, with the logic of subjective expected utility theory' ([2], p. 1682). The irony goes even further.

Corresponding author: Hertwig, R. (sekhertwig@mpib-berlin.mpg.de).

Abnormalities conducive to 'rationality'

In the following sample of investigations, individuals with brain damage or mental illness (henceforth 'abnormal' individuals) were more likely than other ('normal') individuals to adhere to diverse benchmarks of rationality. (i) Patients with damage to the ventromedial prefrontal cortex (VMPFC) were more coherent in their preferences in a consumer choice context (i.e., the Pepsi paradox) [3]. Likewise, they did not fall prey to the correspondence bias and, thus, made more advantageous decisions in an investment context [4]. (ii) In moral dilemmas in which the utilitarian choice (a weaker benchmark of rationality) implies emotionally aversive behavior, patients with VMPFC damage, frontotemporal dementia, or frontal traumatic brain injury showed a greater propensity to make utilitarian judgments [5,6]. (iii) Patients with OFC lesions made choices between gambles that were guided more by the expected value of the gambles than by reported or anticipated regret [7]. (iv) Patients with any of various focal lesions (including damage to the OFC) were less subject to myopic loss aversion and, thus, made more advantageous decisions (resulting in higher income) in an investment task [8]. (v) Participants with a virtual lesion to the right dorsolateral PFC (induced through transcranial magnetic stimulation) exhibited higher acceptance of unfair offers in the ultimatum game [9]. (vi) Patients with autism were less responsive to the framing of monetary outcomes as either losses or gains and, thus, exhibited more internally consistent behavior [10]. A complete list of the studies that we found reporting a positive relation between abnormality and benchmarks of rationality can be obtained by contacting the corresponding author.

What do these results mean?

Before we turn to two issues raised by these findings, a clarification is in order: normal people are not incessant offenders against rationality. By extension, abnormality is not the royal road to rationality, and none of the above authors suggested so. Furthermore, numerous investigations have found, for instance, that lesions to the frontal lobes substantially impair executive functions and that patients with lesions in the VMPFC behave less rationally than do individuals with normal brains (e.g., [11]).

How sane are our benchmarks of rationality?

In light of the observation that abnormality can be conducive to rationality, one may provocatively ask: how sane are various benchmarks of rationality? Only a few of the studies considered in this article explicitly raised this

question [2,6,9]. In a discussion of the indifference of neurological patients to ambiguity versus risk in the urn study cited above, the authors argued: 'Standard decision theory...precludes agents from acting differently in the face of risk and ambiguity. Our results show that this hypothesis is wrong...and suggest a unified treatment of ambiguity and risk as limiting cases of a general system evaluating uncertainty' ([2], p. 1681). Indeed, recent evidence on the description-experience gap supports the conclusion that humans act predictably differently in the face of stated probabilities (risks) and experienced probabilities (representing different levels of uncertainty, depending on the amount of experience [12]). Distinguishing between risk and ambiguity may be not only descriptively correct, but also even advisable on a normative level. Ellsberg stopped short of arriving at this conclusion when he wrote that 'decision-making under uncertainty is still too young to give us confidence that these [Savage] axioms are not abstracting away from vital considerations' ([1], p. 669). One such vital consideration may be that under conditions of ambiguity (uncertainty), the 'brain is alerted to the fact that information is missing, [and] that choices...carry more unknown (and potentially dangerous) consequences' ([2], p. 1683).

Of course, inferring the normative from the normal risks committing a kind of naturalistic fallacy. Yet, normative benchmarks must be open to challenge. For instance, in view of the evidence that *Homo economicus* does not 'get' the tendencies of normal individuals, most of whom will punish unfair behavior in the ultimatum game, even if that means forgoing gains as high as 3 months of income, but endorses those of individuals with a virtual lesion to the dorsolateral PFC [9], the crucial question that some economists have begun to put forth is (see references in [9]): what kind of impoverished notion of rationality are we endorsing that ignores the importance of reciprocal fairness in the fabric of human society and that continues to assume that humans are exclusively self-regarding (with no positive or negative concern for the welfare of others)?

How to measure rationality (or lack thereof)?

The difficulty of investigating rationality empirically is that the behavioral tendencies of individuals with neurological or mental abnormalities can be shown to be disadvantageous, harmless, or even advantageous purely because of the way that experimental stimuli are constructed. For instance, the payoff distributions in the Iowa Gambling Task, often used to study decision-making competence in patients with damage to the VMPFC, can be designed such that a suspected deficit of these patients in reversal learning leads to advantageous rather than disadvantageous choices [13]. Similarly, gambles can be selected such that choosing to minimize anticipated regret (as normal participants did) yields a higher return than does choosing to maximize expected value (as participants with OFC lesions did) [7], although there is no general positive correlation between regret minimization and reward maximization in the gambling domain [14]. Investment environments can be designed such that the propensity of patients with damage to the VMPFC not to become strongly risk-averse (in contrast to normal participants) following successful or failed investments in a previous round can lead to economically more advantageous decisions [8]. Finally, stimuli can be 'rigged up' ([4], p. 1379) so that the causal attributions of patients with damage to the VMPFC result in advantageous investment decisions.

Our key point is this: in designing task environments, experimenters wittingly or unwittingly determine how successful specific behavioral tendencies (whether normal or abnormal) will be. Therefore, experimenters need to be aware of the risk of loading the dice in favor of their own hypothesis. To reduce this risk, experimenters should construct and select stimuli independently of (normal and abnormal) individuals' performance. One important but often neglected yardstick that can inform selection of experimental stimuli is the environment to which the findings are meant to generalize (Box 1).

Box 1. Representative design and an ecological approach to cognition and rationality

Various research programs concerned with human cognition and rationality, ranging from Brunswik's probabilistic functionalism, through Simon's notion of bounded rationality and Anderson's rational analysis, to Gigerenzer and colleagues' notion of ecological rationality (for details, see [15]), assume some form of cognitive adaptation to environmental structures. Given this assumption, it is crucial to consider how stimuli used to measure the performance of individuals and, by extension, their ability to reason rationally, are sampled from the environment. Specifically, stimuli can be sampled from the environment selectively, to demonstrate that the cognitive system can fail (as proof of the existence of violations of rationality), or representatively, to measure the performance of the cognitive system ceteris paribus. Take, for illustration, the overconfidence bias, which is a 'cognitive illusion' in which individuals put too much trust in the accuracy of their knowledge and abilities. Psychologists have often studied overconfidence by presenting participants with general knowledge questions such as: 'Which city has more inhabitants, Atlanta or Baltimore?' Participants select a city and indicate their confidence that the chosen option is correct. Items (pairs of cities) can be systematically selected such that otherwise valid knowledge (probabilistic cues; e.g., Atlanta has a very busy airport whereas Baltimore does not, and cities with a busy airports tend to be more populous) leads to the wrong inference. Alternatively, items can be selected representatively (e.g., randomly) from a predefined reference class (e.g., the largest 200 US cities). A review of 130 overconfidence data sets found that overconfidence was generally pronounced in selected item samples, but close to zero in representative samples [15]. Although the issue of how stimuli are constructed and sampled from the environment is germane for studies reporting that abnormalities can result in more (or less) rational or advantageous decisions, there is a notable lack of concern for it (e.g., [3,4,7]). Representative design also raises a theoretical issue. From the perspective of probabilistic functionalism and ecological rationality [15], the question is not whether a given cognitive process is rational or irrational in itself, but rather in what environments it will succeed or fail. If, for instance, normal individuals become more risk averse after losing or gaining money, whereas abnormal individuals do not, the question is: in which real-world domains is this state-dependent risk-taking strategy adaptive or maladaptive [8]?

Concluding remarks

In conclusion, some recent studies on the link between abnormality and rationality do invite one to challenge the very sanity of some classic norms of rationality. At the same time, they underscore the importance of ecological

Box 2. Emotions take center stage in the link between rationality and abnormality

The studies featured in [2–10] involve patients with either lesions to a range of brain areas or abnormalities such as autism. Despite this heterogeneity, all but one of the studies [2] implicate emotions in the disrupted process that counterintuitively produces more rational behavior. For example, a lack of the 'emotional associations' that are the 'driving force behind...commercial advertisements' enables patients with damage to the VMPFC to show coherent taste-based brand preferences ([3], p. 4). Attenuated 'prosocial sentiments' ([6], p. 614) and weaker emotional reactions to the possibility of causing direct harm to others enable patients with damage to the VMPFC or OFC to overcome emotional revulsion at the 'means' of an action (e.g., smothering a baby to quiet it) and focus on its 'ends' (saving the lives of several others; [5]). Similarly, dampened 'emotional responses' to the possibility of losses liberate patients with 'deficient emotional circuitry' from myopic loss aversion ([8], p. 435, p. 436) and the experience of regret [7]. Humans with damage to the dorsolateral PFC, which weakens 'the emotional impulses associated with fairness goals', are free to follow their selfish impulses without restraint, thereby maximizing material income ([9], p. 829). In addition, a failure to 'integrate emotional contextual cues into the decision making process' enables patients with autism to choose in an internally consistent way, that is, independently of option framing (i.e., loss versus gain) ([10], p. 10746]).

These sketches of underlying processes invite several observations. First, the shared emphasis on emotions suggests that explanations of decision-making and rationality have finally come to recognize the role of emotions in processes once thought to be purely 'cognitive'. Second, these accounts rarely specify the process in question or exactly how emotions impact it. Third, implicit in most of these accounts is the problematic assumption that specific brain regions can be mapped onto emotional processes in a one-to-one fashion; an alternative and, in our view, more accurate, representation is network mapping, in which a given emotion maps onto activity in multiple brain regions and each region takes on different functions at different times. Finally, emotions are cast here as having either a good influence on rationality (e.g., 5–7]) or a bad one [8]. However, ecological rationality suggests that the question is not whether a given emotion (or processes implicating emotions) is good or bad. The core question instead is: in which real-world environments does the emotion result in successful (or unsuccessful) performance?

analysis and representative design. Admittedly, to make a case against any given normative benchmark, it is not enough to demonstrate a link between abnormality and rationality; the demonstration should be complemented by an account of the underlying process. Such an account should first explain the process by which the sane and healthy deviate from a specific norm of rationality. Second, it should (in our view) analyze the performance of this 'intact' process relative to an environment (its ecological rationality or lack thereof). Third, it should explain how disruption of this process (or the brain regions with which the process is thought to be associated) can result in conformity with a given benchmark of rationality. Although none of the studies considered here propose such a complete process account, many offer rough sketches of candidate processes, nearly all of which implicate emotions (Box 2).

Acknowledgments

This research was supported by a grant to R.H. (HE 2768/7-1) from the German Research Foundation (DFG) as part of the priority program on New Frameworks of Rationality (SPP 1516).

References

1 Ellsberg, D. (1961) Risk, ambiguity and the Savage axioms. *Q. J. Econ.* 75, 643–669
2 Hsu, M. *et al.* (2005) Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310, 1680–1683
3 Koenigs, M. and Tranel, D. (2008) Prefrontal cortex damage abolishes brand-cued changes in cola preference. *Soc. Cogn. Affect. Neurosci.* 3, 1–6
4 Koscik, T.R. and Tranel, D. (2013) Abnormal causal attribution leads to advantageous economic decision-making: a neuropsychological approach. *J. Cogn. Neurosci.* 25, 1372–1382
5 Young, L. and Koenigs, M. (2007) Investigating emotion in moral cognition: a review of evidence from functional neuroimaging and neuropsychology. *Br. Med. Bull.* 84, 69–79
6 Mendez, M.F. (2009) The neurobiology of moral behavior: review and neuropsychiatric implications. *CNS Spectr.* 14, 608–620
7 Camille, N. *et al.* (2004) The involvement of the orbitofrontal cortex in the experience of regret. *Science* 304, 1167–1170
8 Shiv, B. *et al.* (2005) Investment behavior and the negative side of emotion. *Psychol. Sci.* 16, 435–439
9 Knoch, D. *et al.* (2006) Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832
10 De Martino *et al.* (2008) Explaining enhanced logical consistency during decision making in autism. *J. Neurosci.* 28, 10746–10750
11 Camille, N. *et al.* (2011) Ventromedial frontal lobe damage disrupts value maximization in humans. *J. Neurosci.* 31, 7527–7532
12 Hertwig, R. and Erev, I. (2009) The description–experience gap in risky choice. *Trends Cogn. Sci.* 13, 517–523
13 Maia, T.V. and McClelland, J.L. (2005) The somatic marker hypothesis: still many questions but no answers. *Trends Cogn. Sci.* 9, 162–164
14 Eagleman, D. (2005) Comment on 'The involvement of the orbitofrontal cortex in the experience of regret'. *Science* 308, 1260
15 Dhami, M. *et al.* (2004) The role of representative design in an ecological approach to cognition. *Psychol. Bull.* 130, 959–988

1364-6613/$ – see front matter
http://dx.doi.org/10.1016/j.tics.2013.08.011 Trends in Cognitive Sciences, November 2013, Vol. 17, No. 11