

# Paradox of Happiness

Ben Eggleston

The paradox of happiness is the puzzling but apparently inescapable fact that regarding happiness as the sole ultimately valuable end or objective, and acting accordingly, often results in less happiness than results from regarding other goods as ultimately valuable (and acting accordingly). That is, in many circumstances, happiness is more effectively achieved when other objectives are regarded as worth pursuing for their own sakes than when happiness alone is regarded as worth pursuing for its own sake (*see* HAPPINESS; HEDONISM; INTRINSIC VALUE). These other objectives might be regarded as ultimately valuable instead of happiness, or merely in addition to happiness; but they must be valued for their own sakes, and not merely as means to the achievement of happiness. These other objectives may include loving family relationships and friendships, meaningful professional relationships, immersion in rewarding work, the exercise of skills and abilities, accomplishments and triumphs, participation in religion or a large cause or movement, and contributions to one's culture or nation.

The paradox of happiness can be understood as applying to people individually or in groups. With respect to people individually, the paradox of happiness is the fact that any given individual is likely to be less happy if happiness is her sole ultimate objective than she would be if other goods were among her ultimate objectives. With respect to groups of people, the paradox of happiness is the fact that any given group of people, such as a community or a society, is likely to be less happy, collectively, if happiness is its sole collective ultimate objective than it would be if other goods were among its collective ultimate objectives.

There are two prominent, mutually compatible explanations for the paradox of happiness. The first appeals to the claim that when people have choices to make, they are generally poor judges of their available options' effects on happiness (both individual and collective). A proponent of this view might claim, for example, that people generally overestimate the extent to which additional income will increase their happiness, and generally underestimate the extent to which a longer commute between home and work will decrease it. In this view, people are systematically so inept at making happiness-promoting choices that a surer route to happiness is for people to aim, when making choices, at objectives other than the promotion of happiness (Sidgwick 1907: 142; Bradley 1927: 102; Parfit 1984: 5).

The second explanation appeals to the claim that even if people were generally capable of accurately judging their available options' effects on happiness, a more subtle problem would remain. This problem is that, for most people, happiness is unattainable without several of the goods listed above (friendship, rewarding work, etc.), and the full realization of these goods typically requires an intensity of

emotional and affective investment – to the extent of regarding the goods as worth pursuing for their own sakes – that is impossible for people who regard happiness alone as worth pursuing for its own sake. According to this explanation, then, people who regard happiness as the sole ultimate objective can expect to be less happy than they might otherwise be, not solely because of the bad choices they will make – though those may be a problem – but also, and more subtly, because of the impoverished range of options and possibilities from which they will have the opportunity to choose (Butler 1726: 186, 190–2; Mill 1873: 117; Sidgwick 1907: 136, 405; Hodgson 1967: 58–9, 61; Stocker 1976: 456; Kavka 1978: 291; Gauthier 1984: 263; Parfit 1984: 6; Railton 1984: 140–1; Elster 1984: 40; Elster 1989: 24).

According to each explanation, the paradox of happiness is ultimately a contingent matter of empirical psychology. Nevertheless, each explanation seems to capture a deep truth about human nature. The paradox they explain has significant, if disputed, implications for ethical theory and other areas of practical philosophy.

In these areas of philosophy, the paradox of happiness is most frequently adduced in order to criticize theories that regard the maximization of happiness – either for the agent alone or for multiple individuals – as the criterion of rational action or right action (depending on the kind of theory being criticized). For example, the orthodox theory of instrumentally rational action maintains that an act is instrumentally rational if and only if it maximally advances the agent's interests (*see* EGOISM). Assuming that a person's interests can be at least roughly equated with happiness, it can then be argued that the paradox of happiness makes this theory self-defeating, in the sense that agents who subscribe to it will be likely to achieve the theory's goal of happiness less effectively than agents who subscribe to a nonorthodox theory of instrumentally rational action (Kavka 1978: 293; Elster 1983: 9; Parfit 1984: 7; Hollis 1998: 17), such as one that regards some of the discrete objectives listed earlier as ultimate ends, instead of regarding the agent's interests or happiness as the sole ultimate end, with other objectives being valuable (if at all) only as means to that end.

The challenge posed by the paradox to this theory of instrumentally rational action is paralleled by challenges posed by the paradox to two ethical theories, ethical egoism and utilitarianism. Ethical egoism maintains that an act is right if and only if it maximally advances the agent's interests, and thus is identical to the theory just discussed except that it is a theory of morality rather than instrumental rationality. As a result, it is vulnerable to the same charge of self-defeat. Similar claims are often made about utilitarianism (*see* UTILITARIANISM; CONSEQUENTIALISM), standard versions of which maintain that an act is right if and only if it maximizes the well-being of all sentient beings (Hodgson 1967: 3, 60; Stocker 1976: 461; Parfit 1984: 27–8). Whereas the charge of self-defeat against the two theories just discussed rests on the individual form of the paradox of happiness, the charge of self-defeat against utilitarianism rests on the collective form of the paradox. Still, the basic contours of the charge are the same: utilitarianism is self-defeating in the sense that agents who subscribe to it will be likely to achieve the theory's goal of maximizing

well-being less effectively than agents who subscribe to an alternative ethical theory such as Kantianism, social contractarianism, or commonsense morality.

In this way, the paradox of happiness makes theories that prescribe the maximization of happiness vulnerable to the charge of being self-defeating. The force of this charge may be contested, however, with a cluster of considerations that advocates of such theories often marshal in defense of them. First, it may be pointed out that for a theory to be self-defeating in the sense employed above is not for the theory to imply its own falsity (which would, admittedly, be damaging). Nor does self-defeat amount to the disproof of any claim of the theory, since theories criticized for being self-defeating often take no position regarding the consequences of people generally subscribing to them. Instead, for a theory to be self-defeating in the sense employed above means something decidedly less discrediting – namely, that, given human beings as they are actually constituted, people who pursue the theory's sole ultimate objective single-mindedly often achieve that objective less effectively than other people. So a theory can be self-defeating because of an accident of human psychology, rather than because of any defect internal to the theory itself. On this view, the paradox of happiness may capture a genuine fact about human nature – one with important practical implications – but it has no bearing on theoretical projects claiming that happiness must, in the last analysis, be acknowledged as the foundation of morality or instrumental rationality (Hare 1981: 38; Scheffler 1982: 46; Kagan 1989: 37).

Although advocates of theories prescribing the maximization of happiness typically maintain that the preceding argument should be an adequate defense against the charge of self-defeat as a matter of principle, they typically add that their theories also respond constructively to the problems that the paradox of happiness causes in practice. For example, given that utilitarianism prescribes the maximization of happiness, it also prescribes that agents shape their personal aims, values, and decision-making procedures in whatever way will best promote the maximization of happiness. Accordingly, utilitarianism may, because of the paradox of happiness, prescribe that agents include things other than happiness among their ultimate objectives. In doing so, it might recommend either of two strategies, depending on which would better promote the maximization of happiness. Each strategy, however, has a corresponding shortcoming, making the choice between them a trade-off.

The first strategy is for agents to keep happiness in mind as their regulative objective that helps them calibrate the intensity of their commitments to other goods that they also regard as ultimate objectives (Hare 1981: 49–52; Railton 1984: 153–4). The main shortcoming of this strategy is that if agents keep happiness in mind as their regulative objective, they might remain vulnerable to the paradox of happiness. For that paradox might plausibly be thought to arise not only in many cases in which happiness is an agent's sole ultimate objective, but also, more generally, in many cases in which happiness is the preeminent value in an agent's decision-making procedure (even if it is not the agent's sole ultimate objective). In light of this consideration, this strategy might plausibly be regarded as missing the point of the paradox of happiness, and thus as equally vulnerable to it (Williams 1988: 190).

The second strategy is for agents to hold goods other than happiness in such high esteem, as ultimate objectives, that happiness, assuming it is still retained at all as an ultimate objective, has no privileged status relative to other ultimate objectives. In short, utilitarianism might be “self-effacing” (Parfit 1984: 23–4, 40–3). The main shortcoming of this strategy is that if agents do not keep happiness in mind as their regulative objective, new complications ensue. One complication is that if a theory prescribing the maximization of happiness directs agents not to regard happiness as having any privileged status in their decision-making procedures – and thus, in effect, directs them to forget the theory altogether – then this seems to open a troublingly wide chasm between the substance of the theory and the kind of thinking it directs agents to engage in. Another complication is more practical. In particular, this strategy is risky to implement, because if an agent decides to revise her values with the aim of making them maximally conducive to happiness, then she faces a difficult task, for there are many ways in which she could do it sub-optimally, but only one way (or perhaps a few ways) in which she could do it optimally. And if the way she chooses involves (as is characteristic of this second strategy) ceasing to regard happiness as a regulative objective, then she has only one chance to get this transformation right. For once she ceases to regard happiness as a regulative objective, then the maximization of happiness will no longer function for her as a goal in the light of which she might monitor, critically assess, and periodically tweak her decision-making procedure. Such ongoing oversight and revision might still occur, but it will be guided by whatever other objectives have gained ascendancy in her value system – not happiness. So this strategy is risky to implement because of the difficulty of getting it right the first time and the probable lack of opportunities for repeated attempts.

Further considerations may be introduced in order to address these objections to the two strategies mentioned, and further objections may then be lodged. Moreover, the paradox of happiness may give rise to other lines of argument in practical philosophy than the ones canvassed here. But these are the lines of argument in which the paradox of happiness has been most influential in historical and contemporary theorizing about ethics.

**See also:** CONSEQUENTIALISM; EGOISM; HAPPINESS; HEDONISM; INTRINSIC VALUE; UTILITARIANISM

## REFERENCES

- Bradley, F. H. 1927. *Ethical Studies*, 2nd ed. Oxford: Clarendon Press.
- Butler, Joseph 1896 [1726]. *The Works of Joseph Butler, D.C.L.*, vol. 2, ed. W. E. Gladstone. Oxford: Clarendon Press.
- Elster, Jon 1983. *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press.
- Elster, Jon 1984. *Ulysses and the Sirens: Studies in Rationality and Irrationality*, rev. ed. Cambridge: Cambridge University Press.

- Elster, Jon 1989. *Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.
- Gauthier, David 1990 [1984]. "The Incomplete Egoist," in David Gauthier (ed.), *Moral Dealing: Contract, Ethics, and Reason*. Ithaca, NY: Cornell University Press, pp. 234–73.
- Hare, R. M. 1981. *Moral Thinking: Its Levels, Method, and Point*. Oxford: Oxford University Press.
- Hodgson, D. H. 1967. *Consequences of Utilitarianism: A Study in Normative Ethics and Legal Theory*. Oxford: Oxford University Press.
- Hollis, Martin 1998. *Trust within Reason*. Cambridge: Cambridge University Press.
- Kagan, Shelly 1989. *The Limits of Morality*. Oxford: Oxford University Press.
- Kavka, Gregory 1978. "Some Paradoxes of Deterrence," *Journal of Philosophy*, vol. 75, pp. 285–302.
- Mill, John Stuart 1981 [1873]. "Autobiography," in John M. Robson and Jack Stillinger (eds.), *Autobiography and Literary Essays*, vol. 1 of *Collected Works of John Stuart Mill*. Toronto: University of Toronto Press, pp. 1–290.
- Parfit, Derek 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Railton, Peter 1984. "Alienation, Consequentialism, and the Demands of Morality," *Philosophy and Public Affairs*, vol. 13, pp. 134–71.
- Scheffler, Samuel 1982. *The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions*. Oxford: Oxford University Press.
- Sidgwick, Henry 1907. *The Methods of Ethics*, 7th ed. London: Macmillan.
- Stocker, Michael 1976. "The Schizophrenia of Modern Ethical Theories," *Journal of Philosophy*, vol. 73, pp. 453–66.
- Williams, Bernard 1988. "The Structure of Hare's Theory," in Douglas Seanor and N. Fotion (eds.), *Hare and Critics: Essays on Moral Thinking*. Oxford: Oxford University Press, pp. 185–96.

## FURTHER READINGS

- Adams, R. M. 1976. "Motive Utilitarianism," *Journal of Philosophy*, vol. 73, pp. 467–81.
- Frey, R. G. 2000. "Act-Utilitarianism," in Hugh LaFollette (ed.), *The Blackwell Guide to Ethical Theory*. Malden: Blackwell, pp. 165–82.
- Hare, R. M. 1989 [1973]. "Principles," in R. M. Hare (ed.), *Essays in Ethical Theory*. Oxford: Oxford University Press, pp. 49–65.
- Pettit, Philip 1988. "The Paradox of Loyalty," *American Philosophical Quarterly*, vol. 25, pp. 163–71.
- Russell, Bertrand 1930. *The Conquest of Happiness*. New York: H. Liveright.
- Woodcock, Scott 2010. "Moral Schizophrenia and the Paradox of Friendship," *Utilitas*, vol. 22, pp. 1–25.