

## Herbert Simon's intelligent altruism

*Saras Sarasvathy*

In a series of published and unpublished exchanges toward the end of the twentieth century, Herbert Simon and Oliver Williamson engaged in arguments about the relationship between TCE and behavioural economics. The exchange derived from personal biography as well as intellectual history and attested to differences not only of ideas but of worldviews. Augier and March (2001) provides one explanation for the disagreement in terms of a family quarrel about the degree to which one ought to assume unbounded rationality, opportunism, and conflict in the creation and sustenance of human organizations.

Another aspect to the quarrel consists in the fact that Simon was not merely or even primarily an economist or social scientist. He was a Renaissance man, a polymath – deeply interested in how the human mind works and how that coheres with how the universe works. Opportunism, like unbounded rationality and conflict, did not make sense to him as fundamentals of human behaviour either empirically or historically and certainly not in terms of what we know about biological evolution. These concepts also did not make sense to him normatively, as guides for future action, for creatures consciously engaged in ‘designing’ their own environments, nor for scientists studying such creatures – scientists of ‘the artificial’ (Simon, 1996).

Concepts such as bounded rationality, docility, and intelligent altruism were more interesting and fundamental to him, even if they were ‘messier’ because (a) they were empirically-based, and (b) they were necessary to hold together a larger, more general tapestry of what we know about human evolution and history within which organizations and markets play smaller roles. Simon explicitly stated that he was not in favour of sacrificing an empirical understanding of how actual humans behaved even if this meant sacrificing simplicity and elegance in the theories that ensued. His argument about Occam's razor is telling in this regard (Simon, 1979, p. 495):

The application of the principle of satisficing to theories is sometimes defended as an application of Occam's Razor. But Occam's Razor has a double edge. Succinctness of statement is not the only measure of a theory's simplicity. Occam understood his rule as

recommending theories that make no more assumptions than necessary to account for the phenomena (*Essentia non sunt multiplicanda praeter necessitatem*).

Throughout a career spanning over six decades, Simon inveighed against economists' penchant for ignoring facts on the ground and clinging to an edifice of economic 'laws' built on increasingly shaky behavioural foundations. In a full-scale argument on this topic in 1997, in a published version of the Raffaele Mattioli Lectures, Simon (1997b, p. 21) emphasized the importance of getting the empirics right:

The [neoclassical] model itself is manipulated with great mathematical formality, and if it is tested with quantitative data, high standards of sophistication are imposed on the statistical methods employed. What is omitted is any serious testing of the validity of the assumptions of the model itself, even the kind of historical, experiential and anecdotal testing that we find in Smith and Marshall. This might be all right if the quantitative, econometric tests were generally sharp and decisive. Almost always, they are not.

The literature of modern economics is full of examples of the sensitivity of models to small changes of assumptions – many, if not most of them beyond the limits of accuracy of statistical tests.

For the most part, Simon has won the argument with regard to bounded rationality, but his challenges continue with regard to other assumptions. One such assumption of particular importance to TCE is opportunism.

### **Opportunism in transaction cost economics**

Just as Simon's early decades were focused on convincing economists and others that 'the major reason why we humans do not behave in a globally rational way was because we cannot,' in the latter decades, he began to focus on the fact that 'we are highly dependent on those around us' (Simon, 1997b, p. 40). Both empirical findings and evolutionary arguments led Simon to conclude that pure self-interest, or its stronger cousin, opportunism, was an inappropriate assumption in models of transactions, not to mention inaccurate and even normatively dangerous for our understanding and design of organizations.

We can trace back TCE's behavioural assumptions about opportunism and ensuing lack of trust in concepts such as altruism, loyalty, and organizational identification, to a partial reading of Adam Smith. For example, economists often cite and build upon *The Wealth of Nations*, (Smith, 1776), in which Smith developed his ideas about the spontaneous benefits of selfishness, an idea he got from Mandeville (1714). The oft-quoted but mostly misunderstood passage from that book concerning the butcher and baker (Baumol, 2002) forms the basis for economists' generalization of the fundamental behavioural assumption about human self-interest. But his (arguably) most important work, *Theory of Moral Sentiments* (Smith, 1759) needs to be consulted if we are to have a complete picture of Smith's position on the matter (Werhane, 1991; Sen 1985). By ignoring this latter work, economists have set the stage for a continuing debate between their own normative theorizing and the observed empirical experience of human transactions, both within and outside organizations.

The notion of opportunism, defined as 'self-interest seeking with guile' is rather recent and can be found in Williamson (1975). It is important to note here that Williamson is not saying that opportunism is always at work; he is only saying that contracts should be made on the assumption that it could be at work. In other words, contracts should be made with an eye to the opportunistic potential they offer (Williamson, 1985).

The work of Douglas North (1978) in economic history illustrates well both the potential for and limits to economic analysis built on assumptions of opportunism. In using opportunism-driven transaction costs as prime movers in economic history, North had to provide an explanation for the observed evidence of successful collective action such as the voting paradox, the absence of free riding in various human organizations, the pervasiveness of charity, and so on. In seeking to explain these manifest phenomena, North introduces the notion of 'ideology' as the missing factor in economic analyses. However, North does not have an underlying explanation why people should buy into 'ideology.' Similarly, scholars who argue in a more strictly neoclassical vein, such as Stigler and Becker (1977), Becker (1996), and Gauthier (1984) all have the problem of being unable to explain why rational actors would acquire beliefs and behaviours that are not in their narrow self-interest. While North's conclusions also cohere with those of Arrow (1974) and Sen (1985), others have posited other exogenous factors such as the Protestant ethic (Weber, 1905).

## **Docility and intelligent altruism**

Simon has a different explanation, a behavioural construct that brings together experimental evidence and formal evolutionary models to provide an alternative assumption for transactional and organizational economics. That construct is *docility*. Simon defined it as: *The tendency to depend on suggestions, recommendation, persuasion, and information obtained through social channels as a major basis of choice.* (1993, p. 156) Docility follows directly from the limitations of human cognition. Through a series of articles, essays and lectures, Simon developed a notion of ‘intelligent’ altruism based on this notion of docility to argue that bounded rationality not only limits our ability to undertake the computational demands of highly opportunistic behaviour, but also selects such behaviour out (in an evolutionary sense) and selects in those who are willing and able to depend on others and help sustain others in a group – i.e. intelligent altruists (Simon, 1997a and b).

More recently, Knudsen (2003) has argued for the role of docility in the emergence of altruism in biological populations. The case for the evolutionary dominance of intelligent altruists is also well-argued from perspectives other than those resting on docility. Hill (1990) for example shows that under the normal assumptions of neoclassical economics, the invisible hand of the market will tend to weed out persistently opportunistic behaviour. Interestingly enough, without resorting to evolutionary arguments, Adam Smith himself had made the case for the fundamental behavioural assumption of persuasion in all economic exchanges:

Different genius is not the foundation of this disposition to barter which is the cause of the division of labour. The real foundation of it is that principle to persuade which so much prevails in human nature... We ought then to mainly cultivate the power to persuasion, and indeed we do so without intending it. Since the whole life is spent in the exercise of it, a ready method of bargaining with each other must undoubtedly be attained. (Smith, 1978, pp. 493-494)

Game-theoretic and behavioural-economic evidence of all kinds has been accumulating over the last two or more decades pressing us to rethink the assumption of pure opportunism. In fact, what we know about self-interest based on empirical evidence, both in the lab (See Rabin, 1998 for a comprehensive review) and in the field suggests that neither opportunism nor altruism or trust can possibly form clear bases for predictions about human behaviour. Both are confounded

not only by heterogeneity in behavioural traits and choices, but are situated and change over time. Dawes and Thaler (1988) capture this in an eloquent passage:

In the rural areas around Ithaca it is common for farmers to put some fresh produce on the table by the road. There is a cash box on the table, and customers are expected to put money in the box in return for the vegetables they take. The box has just a small slit, so money can only be put in, not taken out. Also, the box is attached to the table, so no one can (easily?) make off with the money. We think that the farmers have just about the right model of human nature. They feel that enough people will volunteer to pay for the fresh corn to make it worthwhile to put it out there. The farmers also know that if it were easy enough to take the money, someone would do so.

Besides the negative evidence against any default propensity for opportunism, there is also direct positive evidence for the construct of docility -- the fact that human beings are prone both to give and take advice. In a summary of findings on the subject, Schotter (2003, p. 196) concludes:

- (i) Laboratory subjects tend to follow the advice of naïve advisors (i.e. advisors who are hardly more expert in the task at hand than they are).
- (ii) This advice changes their behavior in the sense that subjects who play games or make decisions with naïve advice play differently than those who play identical games without such advice.
- (iii) The decision made in games played with naïve advice are closer to the predictions of economic theory than those made without it.
- (iv) If given a choice between getting advice or the information upon which that advice was based, subjects tend to opt for the advice, indicating a kind of underconfidence in their decision-making abilities that is counter to the usual egocentric bias or overconfidence observed by psychologists.
- (v) The reason why advice increases efficiency or rationality is that the process of giving or receiving advice forces decision-makers to think about the problem they are facing in a way different from the way they would if no advice were offered.

For Simon the important point was that most of what we do we have learned from those around us because

[b]ehaving in this fashion contributes heavily to our fitness because... the information on which this advice is based is far better than the information we could gather independently. As a result, *people exhibit a very large measure of docility*. (Simon, 1993, p.157, italics in original).

By fitness, Simon meant our biological fitness i.e. the effectiveness with which we produced progeny. Simon theorized that social information contributes so heavily to our fitness – it is such an adaptive trait - that docile human beings drove out non-docile human beings *in actual evolutionary competition*. He believed that

[t]he farther the complexities of the real world extend beyond our capabilities for knowledge and calculation, the more valuable is docility, to enable us to benefit from the collective knowledge and skill of our society. (Simon, 1997, p. 41).

From a gene's-eye perspective, docility would be strongly selected for because it enhances the reproductive fitness of each human being. In other words, docility is the individual's way of leveraging the advantages of social living (including the rudimentary division of labour, which is thought to extend a million years into human history: Ridley 1995). As a result of being fitness-enhancing at the level of the gene, the population of human beings that now walks the planet exhibits a large degree of docility.

It is interesting to speculate what a TCE that took docility as its fundamental behavioural assumption would look like. Perhaps we would need to go beyond bargaining to negotiation – and not merely distributive negotiation (aimed at dividing the pie) but integrative negotiation to grow the pie under negotiation (Pruitt, 1981; Pruitt and Lewis, 1975; Pruitt and Rubin, 1986). In other words, TCE could be incorporated into growth models and in the economics of innovation in new ways – not only competition, but cooperation could also serve as a discovery procedure (Hayek, 1945; 1984). In fact, the results summarized by Schotter (2003) suggest an explicit role for docility in coming up with creative solutions to problems. Similarly docility-based TCE might provide ways to move beyond static models of social networks to dynamic ones. In a worldview in which opportunism is not such a big hurdle and the ability and willingness to

persuade and be persuaded is a dominant characteristic, humans will be able to build new relationships more easily and grow their networks faster. This could bring to the fore interesting challenges to current static models such as Burt's (1992) structural holes.

### **Simon the man**

It is perhaps appropriate to conclude a biographical sketch with the scholar's personal stance on the subject of behavioural assumptions at the heart of our choices in building a better society. In terms of personal political preference, Simon was a New Deal Democrat. In a characteristically profound yet simple way, he explained his choice as follows (Simon, 1991, p. 133):

The reason is depressingly simple, and has little to do with the wisdom or unwisdom of specific policies of either political party. Among the fundamental problems in every society, two stand out. People have to be motivated to contribute to the society, to produce. At the same time, they have to be protected if they are unable to take care of themselves adequately. You can think of it as the balance between incentives and distributive justice. Too much concern with the latter may weaken the former, and vice versa.

Using this simple-minded dichotomy, you can classify people (roughly) into two groups by their answers to the following question: Is it more important that (a) all chiselers be detected and removed from the welfare lists, or (b) no sparrow should fall from Heaven unseen and uncared for. If the answer is (a), the respondent is a Republican; if (b), a Democrat. Either answer is rationally defensible. I just happen to prefer the second one.

In a world where 'liberal' has become a bad word that current American politics is striving to re-define, it is perhaps time for economists too to rethink the fundamental assumption about human behaviour that is used to explain and inform the design of our social and political institutions. It is cause for comfort to know that while Simon, the scientist, urged this from an empirical standpoint, Simon, the person, also believed it would be a good guiding principle for building a better society.

### **References**

Arrow, K. J. (1974), 'Limited Knowledge and Economic Theory', *American Economic Review*, 54: 1-10.

- Augier, M. and J. G. March (2001), 'Conflict of Interest in Theories of Organization: Herbert A. Simon and Oliver E. Williamson', *Journal of Management and Governance*, 5(3-4): 223-230.
- Baumol, W. J. (2002), *The Free-Market Innovation Machine: Analyzing the Growth Miracle of Capitalism*, Princeton: Princeton University Press.
- Burt, R. (1992), *Structural holes: The social structure of competition*, Harvard University Press.
- Dawes, R. M., & R. H. Thaler (1988), 'Anomalies: Cooperation', *The Journal of Economic Perspectives* 2: 187-197.
- Gauthier, D. (1984), 'Deterrence, Maximization, and Rationality', *Ethics* 94: 474-495.
- Hayek, F. A. (1945), 'The Use of Knowledge in Society', *American Economic Review*, 35(4): 519-30
- Hayek, F. A. (1984), 'Competition as a discovery procedure', in C. Nishiyama and K. Leube, eds., *The essence of Hayek*. Stanford: Stanford University Press, p. 257.
- Hill, C. W. L. (1990), 'Cooperation, Opportunism, and the Invisible Hand: Implications for Transaction Cost Theory', *Academy of Management Review*, 15: 500-514.
- Knudsen, T. (2003), 'Simon's Selection Theory: Why Docility Evolves to Breed Successful Altruism', *Journal of Economic Psychology*, 24: 229-244.
- Mandeville, B. (1714), *The Grumbling Hive* (re-issued, with notes, as *The Fable of the Bees*).
- North, D. C. 1978, 'Structure and Performance: The Task of Economic History', *Journal of Economic Literature* 16: 963-978.
- Pruitt, D. G. (1981), *Negotiation Behavior*, New York: Academic Press.
- Pruitt, D. G. and S. A. Lewis (1975), 'Development of integrative solutions in bilateral negotiation', *Journal of Personality and Social Psychology*, 31: 621-633
- Pruitt, D. G. and J. Z. Rubin (1986), *Social Conflict: Escalation, Stalemate, and Settlement*, New York: Random House.
- Rabin, M. (1998), 'Psychology and Economics', *Journal of Economic Literature* 36: 11-46.
- Schotter, A. (2003), 'Decision Making with Naïve Advice', *American Economic Review* 93(2): 196-201.
- Sen, A. (1985), 'Well-Being, Agency and Freedom: The Dewey Lectures 1984', *The Journal of Philosophy* 82: 169-221.

- Simon, H. A. (1979), 'Rational Decision Making in Business Organizations', *American Economic Review* 69(4): 493-513.
- Simon, H. A. (1991), *Models of My Life*, New York, NY: Basic Books.
- Simon, H. A. (1993), 'Altruism and Economics', *American Economic Review* 83: 156-161.
- Simon, H. A. (1996), *Sciences of the Artificial*, Cambridge: MIT Press.
- Simon, H. A. (1997a), 'A Mechanism for Social Selection and Successful Altruism', *Models of bounded rationality*, 3: 205-16.
- Simon, H. A. (1997b), *An Empirically Based Microeconomics*, Cambridge, UK: Cambridge University Press.
- Smith, A. (2000 [1759]), *The Theory of Moral Sentiments*, *Great Books in Philosophy*, Amherst: Prometheus Books.
- Smith, A. (2002 [1776]), *An Inquiry into the Nature and Causes of the Wealth of Nations*, Cambridge: IndyPublish.com
- Stigler, G. J., & G. S. Becker (1977), 'De Gustibus Non Est Disputandum', *American Economic Review* 67: 76-90.
- Weber, M. (2001 [1905]) *The Protestant Ethic and the Spirit of Capitalism*, *Classics Series*, New York: Routledge.
- Werhane, P. (1991), *Ethics and Economics: The Legacy of Adam Smith for Modern Capitalism*, Oxford University Press.
- Williamson, O. E. (1985), *The Economic Institutions of Capitalism, Firms, Markets Relational Contracting*, New York: Free Press.
- Williamson, O. E., M. L. Wachter and J. E. Harris (1975), 'Understanding the Employment Relation: The Analysis of Idiosyncratic Exchange', *The Bell Journal of Economics* 6: 250-278.