Simon's Selection Theory

Why Docility Evolves to Breed Successful Altruism

Knudsen, Thorbjørn

Simon's Selection Theory:
Why Docility Evolves to Breed Successful Altruism*

By Thorbjørn Knudsen

Department of Marketing
University of Southern Denmark
Campusvej 55
DK-5230 Odense M
Denmark

tok@sam.sdu.dk

For the special issue of the Journal of Economic Psychology in honour of Herbert A. Simon

Final revision of August 2002

Word count: 7.600 words including footnotes and references.

Keywords: Altruism, Bounded Rationality, Evolutionary Economics, Natural Selection, Organisational Behaviour.

PsycINFO classification: 2340, 2360, 2910, 3020, 3660.

JEL classification: D23, D64, D83, B52, M10.

* The author appreciates helpful suggestions from Meta Andrés, Mie Augier and James G. March and an anonymous reviewer. All remaining errors were produced without any help.

Simon's Selection Theory:

Why Docility Evolves to Breed Successful Altruism

Abstract

In light of the under-explored potential of Simon's theory of altruism, the purpose of the present article is to review his explanation of altruism and to point out some of its implications for behavioural economics and theories of economic organization. In the course of the argument, this article relates Simon's theory of altruism to Hamilton's theory of kinship selection and then proceeds to examine a critical assumption of Simon's model that social organizations know better than individuals. Within the parameters of Simon's own model, the paper suggests how this assumption can be justified. The paper concludes by noting that Simon offered a new and so far under-explored mechanism for the emergence of altruism in biological populations and suggests a controlled experiment to test Simon's explanation against Hamilton's. Finally, it is noteworthy that Simon's theory has immediate implications for the understanding of human nature that invites revision and development of behavioural economics and theories of economic organization.

Simon's Selection Theory:

Why Docility Evolves to Breed Successful Altruism

Although altruism seems a ubiquitous feature of nature and society, it is hard to explain how it can be sustained when defined as providing a benefit for others at a cost to the provider. From the viewpoint of evolutionary theory, altruism carries the cost of a reduction in the altruist's fitness.[1] That is, the altruist will on average have less offspring than the selfish organism. Therefore, the population share of altruistic organisms will eventually vanish, and the selfish ones will prevail. If altruism is a hardwired trait, the genetic configuration that codes for it will cease to exist, and if it is a learned trait, unlearning will set in. The logic appears flawless, and yet altruism is thriving in nature and society. Why? Assuming altruism to be a hardwired trait, kinship selection was the convincing mechanism identified by Hamilton (1964). Close kin to a high degree share genes. It is therefore possible that the fitness reduction of altruism is less than the fitness increase weighted by the coefficient of relatedness. As relatedness increases, the probability of the sustained presence of altruism in a population also increases.

Hamilton's (1964) formulation was later extended to encompass clustering in geographical space as a general condition favouring the emergence of altruism (Bergstrom, 2002; Hamilton, 1975; Myerson, Pollack & Swinkels, 1991). The effect of the tendency to cluster, termed assortment or viscosity, is that altruism may evolve if the population is genetically related or spatially clustered. Hamilton's rule, that the cost of altruism must be less than the benefit weighted by relatedness, is thought to be a general hurdle for altruism

[1] Simon uses the term "altruism" in a technical sense that must be distinguished from its broader meaning in ordinary parlance as the unselfish concern for the welfare of others. Thus Simon defines altruism according to its use in evolutionary theory (genetics) as forgoing progeny to the benefit of others. For example, Simon (1990, p. 1665) defines altruism as "behavior that increases, on average, the reproductive fitness of others at the expense of the fitness of the altruist."

that can be surpassed only under rather specific circumstances (Axelrod, 1997; Bergstrom, 2002; Knudsen, 2002a; Myerson et al., 1991; Sella & Lachman, 2000). Thus, an alternative explanation for the viability of altruism that differs from both kin-selection and geographical clustering is a repeated game in which agents may develop reciprocal altruism (Axelrod, 1997). If a repeated game ends with a probability less than one, altruism may survive, even if this would be impossible, were it known at which time step the game ended.

As indicated by the amount of effort allocated to develop convincing models that can explain the viability of altruism, it is an important issue in understanding observed biological, social and economic behaviour. It is therefore interesting that Simon in a stream of publications (Simon, 1983; 1990; 1991; 1993; 1997) offered a convincing explanation for the viability of altruism that differed from those mentioned above. It is here important to note that Simon's model is defined on the basis of an "even if" argument that makes it applicable to both economics and biology. Even if genes are the controlling sites of natural selection, Simon's model shows that the general capacity to learn at the socio-economic level may cause positive selection of altruism. As discussed below, the implication is that Simon's model deserves further attention both in economics and biology.

In light of the under-explored potential of Simon's theory of altruism, the purpose of the present article is to review his explanation of altruism and to point out some of its implications for behavioural economics and theories of economic organization. In the course of the argument I relate Simon's theory of altruism to Hamilton's (1964; 1975) theory of kinship selection and structured populations and then proceed to examine a critical assumption of Simon's model that organizations know better than individuals. Within the parameters of Simon's own model, it is suggested how this assumption can be justified.

How altruism may spring from the capacity to be instructed

Simon's model of altruism was based on the two assumptions of bounded rationality and docility, both introduced as empirical assumptions to characterise fundamentally the human condition (Simon, 1947; 1955; 1956; 1979; 1987).

Bounded rationality refers to limits of knowledge and computational power of individuals, and the implications of these limits in terms of the possible alternatives of an action that can be known, and the possibility of evaluating the consequences of each alternative. For instance, people consider treatments for physical symptoms of illness by including some but not all possible treatments. And in the course of evaluating a set of treatments, the screening of information is imperfect in the sense that a good treatment may be rejected, and a bad treatment accepted. Or the uncertainty regarding the consequences of alternative treatments may be so high that the choice is based on a procedure that is akin to the flipping of a coin.

Docility refers to the capacity for being instructed and the tendency to accept and believe instructions received through social channels. In an evolutionary perspective, "docility" must be seen as an evolved capacity expressed as "the tendency to depend on suggestions, recommendations, persuasion, and information obtained through social channels as a major basis for choice" (Simon, 1997: p. 244). It must be noted here that Simon was not entirely satisfied with the term docility because of its passive overtones and would have preferred the term "socializability", were it not so awkward (Simon, 1997: p. 3). Rather than being passive, Simon viewed a docile person as one that readily learned in a social setting and tended to acquire socially approved behaviours and beliefs.

Simon's definition of docility encompassed a cognitive and a motivational component. The cognitive component referred to the tendency to form beliefs on the basis of information received from legitimate or "qualified" sources rather than relying on a personal

evaluation. For example, in evaluating a treatment for a physical symptom of illness a docile person would tend to rely on advice from a certified professional member of a medical institution rather than cooking up a treatment on the basis of personal studies. Today, the docile person would on average experience a great advantage over the non-docile by tapping in to the cumulated wisdom of the medical profession. By the same token, in a different time and social context, a docile person would tend to accept the treatment prescribed by a witchdoctor. Perhaps the advantage of docility in this case lies more in the comfort of tapping in to a well-known set of values than basing the cure on objective evidence. This possibility is an expression of the motivational component of docility, to accept information on the basis of social approval rather than individually held motives that are not socially acquired.

With bounded rationality and docility in place, the road is cleared for Simon's evolutionary explanation of altruism to unfold. Consider a population composed of p altruists and $1-p$ selfish individuals, and let X be the fitness of a population considered within their current environment. Define F_S as the fitness of selfish individuals and F_A as the fitness of altruists. By definition, the typical altruist will sacrifice c units of offspring to contribute b units of offspring to the population (including self). The fitness functions are:

(1) $F_S = X + bp$

(2) $F_A = X + bp - c$

It is clear from these fitness functions that the altruists suffer a comparative disadvantage of c and will eventually vanish.

At this point it is interesting to briefly visit Hamilton's (1975) model of structured populations. Hamilton (1975) used the Price (1970; 1972) equations to introduce an assortment procedure so the correlation in two randomly selected members of a population is

F. According to Hamilton, the positive correlation *F*, also referred to as the viscosity parameter, can be interpreted as the chance that the *b* units of fitness are definitely given to a fellow altruist; while with chance *(1-F)* they are given (as they always were) to a random member of the population. Then, if the altruist gives up *c* units of fitness to add *b* units to joint fitness, the above fitness functions become:

(3) $F_S = X + bp(1-F)$

(4) $F_A = X + bp(1-F) + bF - c$

This implies that the criterion for positive selection of altruism, known as Hamilton's rule, is:

(5) $bF > c \Leftrightarrow b/c > 1/F$

Hamilton's rule is general for asexual models with non-overlapping generations, and it also holds for diploid biological models. In other words, it is general to both social and biological populations. The main conclusion to be drawn from Hamilton's model is that altruism or costly cooperation can evolve only if the population is genetically related or spatially clustered so the viscosity parameter $F > 0$.

Since Hamilton's rule is commonly thought to be a general hurdle for altruism to evolve, it is interesting to note that Simon devised a simple alternative mechanism that works, even if Hamilton's rule is not obeyed. That is, if a population is completely homogenous, so $F=0$. For this reason Simon's model deserves much more attention than it has so far received.

Invoking the assumption of bounded rationality, Simon argued that human beings and other creatures do not behave optimally for their fitness because they do not possess the knowledge and computational power required to support optimisation. In the terminology

introduced above, people are only capable of considering a limited subset of the possible actions that may bear on their fitness and are, at best, only capable of an imperfect screening of information regarding the consequences of each action.

Because of bounded rationality, people (and organisms) will on average benefit from receiving and relying on the information received through social channels. Unless society and its organizations generally cumulate bad information, docile people who enjoy a capacity to be instructed by society and to acquire skills on the basis of this socially transmitted information will experience a fitness advantage over those lacking the capacity. Thus, according to Simon, docile individuals experience a fitness advantage of *d* units over non-docile people because of their use of socially transmitted skills and their conformity to socially sanctioned behaviour. This fitness advantage may now be "taxed" by society through influencing the docile people to engage in behaviours that are useful to society in the sense that they contribute to net fitness. If society imposes a "tax" on docility by prescribing behaviours that reduce fitness by *c* units, the fitness functions for a homogeneous population are:

(6) $F_S = X + bp$

(7) $F_A = X + bp + (d-c)$

Note first that the "tax" *c* contributes to the net fitness of society, and therefore it is a measurement of altruism. Note further that altruism will dominate selfish behaviour as long as the fitness advantage from docility is larger than the fitness reducing "tax" imposed by society, i.e. whenever $d > c$. But why do smart altruists accept instructions indiscriminately? Why do they not avoid the contribution *c* to society? According to Simon, there are two reasons for this. First, referring to the assumption of bounded rationality, even smart people

have too limited knowledge and computational power to screen off the fitness decreasing instructions received through social channels. Second, the tendency to conform to social norms and accept information on the basis of social approval will reinforce the individual's limited ability to discriminate between fitness increasing and fitness decreasing instructions. This completes Simon's model of altruism.

In summary, altruism is in this model a by-product of docility. Because of bounded rationality, a docile person must indiscriminately accept the fitness advantage *d* and the fitness reducing cost *c*. And whenever *d* is larger than *c*, there is positive selection for docility. In Simon's model, altruism springs from docility, an evolved capacity to be instructed. A further force supporting the positive selection of docility is the emergence of social sanctions curbing deviant behaviour. Altruism therefore includes norms that encourage time spent on instructing others in proper social behaviours as well as enforcing proper behaviours by encouragement, detection and punishment of deviant behaviour.

One assumption of Simon's model, that society and its organizations know better than individuals, however, invites to further consideration. If we use Simon's model and by implication assume, it is valid also for social organizations, the question will be whether it is reasonable to infer that organizations know better than individuals? If this were not the case, the term *d* becomes negative since the docile individual would receive fitness reducing instructions through social channels. There would thus be selection against docility. Although this problem seems irrelevant in view of today's advances in the accumulation of societal knowledge, it must be remembered that Simon's argument relies on the possibility that the capacity for docility was positively selected also in ancient times when "society" was no more than a band of hunters and gatherers. In this case, as in the case of modern social organizations, the question whether an organization of *N* individuals better screens the distribution of prospects on offer than the sum of *N* independent individuals must be

answered in the affirmative. To further justify Simon's model of altruism, it is therefore necessary to show that organizations, at least in some instances, better screen information than the members would do as independent agents. This is done in the following.

Before turning to this problem, Hamilton's model of altruism is revisited and integrated with Simon's. The point is to show that in the case of a clustered population there may be positive selection for altruism, even if society imposes a fitness reducing tax *c* that is higher than the fitness advantage *d* associated with docility. By considering the possibility that a population may be clustered, as measured by the viscosity parameter *F*, we get:

(8) $F_S = X + bp(1-F)$

(9) $F_A = X + bp(1-F) + bF + (d-c)$

In a clustered population, altruism dominates selfish behaviour whenever $d > c-bF$. Thus altruism may evolve in a clustered society imposing a "tax" that is almost *bF* units of fitness larger than the benefit *d* of docility. According to Simon's model, we should expect altruism to be a pervasive feature of every society, clustered or not. As the fitness advantage of docility increases by society's accumulation of useful empirically tested knowledge and processes that further utilize the members' capacity to be instructed, the term *d* will grow and the possibility of a correspondingly large term *c* arises. As Simon (1997: p. 246) notes, there is indeed room for a lot of altruism in a modern high-tech society. But also in a low-tech society, some level of altruism is possible despite a negative contribution of docility. In early times, there was room for some level of docility, even if the members of a social group on average received bad advice. In particular, as long as the docile person would tend to team up with other docile people, and the less docile person tends to wander off, altruism would evolve, but only slowly. As societies grew increasingly homogenous and accumulated

increasingly more sound knowledge, a strong positive selection for docility would eventually occur. In this explanation, bounded rationality may be seen as a universal primordial feature characterising every being, whereas docility is a capacity that may evolve, given the right conditions. The primary condition favouring the evolution of docility is the appearance of (social) organisations that accumulate increasingly useful knowledge. To be convincing, Simon's explanation of altruism must therefore ideally be supported by an argument showing that the pure structural features of social organisations, independent of any form of technological advances, can lead to better screening of the prospects offered by the environment. In this case, there would be selection for docility, even in the ancient times when the media of storing knowledge were rudimentary or limited to primitive behavioural programmes.

Altruism in Social Organizations

In a summary of his work on altruism, Simon (1997) notes that people in recent theories of economic organization (the new institutional economics) are solely motivated by economic gains while the empirical facts apparently also include motives such as organizational loyalty and identification. Simon (1997: p. 199) further notes, "[w]orking my way through this conundrum has been the greatest challenge in my economic research in the past two decades." As we have seen above, one outcome of this work is a model of successful altruism to complement Hamilton's and its recent extensions. Another outcome, to be the focus of this section, is that docility obviously may be at work at the level of social organizations and therefore support the evolution of extensive altruism also in modern business organizations.

Why do employees often work hard? Arguing that purely economic rewards only provide some of the explanation, Simon (1991; 1993; 1997) identified a number of alternative mechanisms that may secure adherence to organizational goals. The most important is

organizational identification. It is an empirical fact, Simon (1991; 1993; 1997) argued, that society is permeated by loyalties to various groups, including ethnic, religious and family groups and that such loyalties provide a crucial basis for altruism.

As noted above, the docility argument can be extended to any form of social group. When there is positive selection for docility, the member of any social group will tend to accept the instructions transmitted within the group. These quite possibly include instructions serving to inculcate organizational loyalties among its members and to further the members' identification with the goals of the organization. As Simon (1991; 1993; 1997) noted, such instructions that serve to build organizational identification will have a tendency to clearly discriminate between the members of the organization and the non-members.

From the viewpoint also of the modern business organization, the promotion of organizational identification by instilling organizational pride, loyalty and values therefore has an obvious advantage in motivating the employees to actively pursue the goals of the organization (Simon, 1991; 1993; 1997). For the individual employee, Simon's (1991; 1993; 1997) argument established that pride, loyalty, and values associated with membership of an organization are important motivating factors along with material rewards.

One very important implication of this argument is that it helps explain why an employee would not consistently act in economic self-interest but often attempt to aim at realising the goals of the organization, e.g. to maximize organizational profits. As Simon (1991: 1993: 1995) repeatedly emphasized, contrary to the assumptions adopted by the new institutional economics, a theory that includes aspects of motivation should begin with empirically valid postulates about real people in real organizations. In particular, he argued that assuming employees will inadvertently maximize their personal utility by shirking and cheating seems unfounded in the face of the available evidence supporting the competing argument that employees may indeed adopt organizational goals as their own. Yet this

implies that employees may actually accept instructions at a personal disadvantage. Why would this be the case? Why does the individual employee not sharply discriminate between the instructions providing personal advantage (in terms of wealth or utility) and those providing a disadvantage?

First, because of the assumption of bounded rationality, the individual is incapable of perfectly screening the consequences associated with each instruction. Second, according to Simon's selection argument, people will generally tend to be docile and therefore accept the instructions received through social channels. The typical employee will therefore tend to be docile and to accept most instructions transmitted within the organization. Also, the possibility that the individual may limit the set of the possible goals to include only those of the organization (or organizational subunit) offers the employee a convenient means to simplify an otherwise more complex situation. Note here that Simon's (1991; 1993; 1997) argument implies that over time the employee will increasingly rely on the instructions received within the membership organization and increasingly ignore instructions and information transmitted from alternative sources.

This opens the question whether an organization of *N* individuals better screen the distribution of prospects on offer than the sum of *N* independent individuals. If this was not the case, it seems possible that the tendency to develop organizational identification would eventually weaken, because independent producers and traders would consistently outperform organizations. In the following, it is shown by example that organizations, at least in some instances, better screen information than the members would as independent agents. Since this possibility remains and since it would be very hard for the individual to infer what kind of organizational structures actually help make better decisions, we conclude that Simon's model of altruism rests on firm ground as an explanation of the emergence of altruism in any form of social organization.

Why Social Organizations Sometimes Know Better than Individuals

If a collection of independent individuals were always better off compared to a social organization including these individuals as members, the docility term in Simon's model would be negative and altruism could not evolve. It is therefore critical to show that, at least in some instances, the social organization helps structure the information flows among individuals to improve their ability to screen the distribution of prospects currently offered by the environment. Here, a prospect may refer to any opportunity that de- or increases the agent's benefit by whatever dimension(s) of benefit we wish to analyse. To keep things simple and consistent with Simon's model of altruism, we define the distribution of offers in terms of fitness (expected number of offspring). In the following, some necessary terminology is first introduced. Then comes the argument supporting Simon's model of altruism.

Within the parameters of Simon's model, bounded rationality implies that individuals are at best only capable of imperfectly screening a distribution of prospects. That is, individuals sometimes reject good offers (Type I error) and accept bad offers (Type II error). A useful way to model this problem is to impute to the individual a screening function that maps the distribution of offers onto the probability of accepting an offer. To do this, the distribution of offers is indexed with respect to its consequences (e.g. fitness, profit or utility) and the probabilities are constrained within the closed interval *[0,1]*. Given a distribution of offers, the condition of bounded rationality may be modelled as a screening function *f* mapping the offer *x* onto the probability *p* that an individual accepts the offer, so *p=f(x)*. The more severe the condition of bounded rationality, the larger the deviation from the perfect screening function, i.e. the higher the incidence of Type I and Type II error. In the most

severe case, the agent simply has no clue about the distribution of prospects and picks an arbitrary constant $K \in [0,1]$, and the screening function becomes $K=f(x)$.

In order to proceed, it is useful to conceptualise the structure of a social organization as a connected finite graph *G* including all possible paths of traversal of prospects among its members. The traversal of prospects can be defined in terms of a dynamic rule that determines where a member can or must send a prospect when it is accepted or rejected. A prospect can either be finally rejected or accepted. In the case of acceptance, the prospect adds to the accumulated knowledge of the organization. This can be symbolized by including the prospect in the organization's "final portfolio."

Given this terminology, the ability of the social organization to support its members' screening of prospects can be defined in terms of the graph screening function *G* and the probability *P* (not necessarily a "hard" probability) that the social organization will accept a prospect *x*, so $P=G(f(x))$. An analytical expression of *G* can always be derived by recursively traversing all the possible acceptance and rejection paths among the agents of a structure (Christensen & Knudsen, 2002a).

The simplest structures of some interest, a serial and a parallel connection of two agents were analysed by Sah & Stiglitz (1986, 1988). Intimately connected with the structure of two serially connected agents, referred to as a "hierarchy," Sah & Stiglitz (1986, 1988) defined a dynamic rule according to which the second agent only received the prospect if the first agent accepted it. Or the other way around, in a serial connection, any agent has the authority to reject a prospect. A parallel connection, referred to as a "polyarchy," was defined in the following way: With probability 0.5 one of two agents first receives a prospect. If this agent chooses to accept, it is accepted by the social organization and included in its final portfolio. If, on the other hand, this agent chooses to reject it, the second agent receives the prospect. If the second agent accepts the prospect, it is accepted by the social organization

and included in its final portfolio. If the second agent, however, also chooses to reject it, this prospect is rejected for good and dumped by the organization.

Note here that the dynamic rule imputed to the polyarchy structure cannot be inferred by the structure itself, a difficulty preventing any straightforward generalisation of this structure defined by Sah & Stiglitz (1986). As Christensen & Knudsen (2002a) note, it is necessary to define a rule for polyarchy-rejection and -acceptance in order to generalise the polyarchy to structures including three or more agents.

For the current purpose, the important point is that Sah & Stiglitz (1986; 1988) showed that a hierarchy is the best performing structure in bad times, defined in terms of an expected value of the distribution of prospects less than zero. In good times, the polyarchy is the best performing structure. Since Sah & Stiglitz (1986; 1988) omit a comparison between the individual agent and the structure and since their analysis is limited to explicit structures including only two agents, it would be important in order to support Simon's docility argument if at least some large structures including three or more agents could be shown to perform better than the members would as independent agents. As Christensen & Knudsen (2002b) show, it is indeed possible that any graph (including deterministic and stochastic graphs as well as graphs with feedback) for some cost models may outperform a collection of independent agents equal in number to those employed in a social organization.

As indicated above, it is useful for the purpose of analysing screening functions to conceptualise a social organization as a graph. In order to generalise the prospect evaluation model, it is further useful to distinguish between stochastic and deterministic graphs. In a stochastic graph the receiving agent is picked at random by the sender. Given a dynamic rule, the receiver may be fixed, however, in which case the graph is called outward deterministic; or the sender may be fixed, in which case the graph is called inward deterministic. A stochastic graph thus represents a completely unstructured social organization, whereas a

deterministic graph (both inward and outward) represents a social organization with the highest degree of internal structure. From the viewpoint of modelling real word social organizations, deterministic graphs seem the most relevant. Since employing meta-agents in nested structures easily reduces stochastic graphs, including three or more agents, to a manageable size, deterministic graphs also appear most interesting from a technical viewpoint (Christensen & Knudsen, 2002a).

Insert Fig. 1 about here

The following example, Fig. 1 above, shows a deterministic graph in which information always enters the social organization at the coordinate (1,1). In the case of serial processing, a dynamic rule is used that generalizes Sah & Stiglitz's (1986) accept and reject rule for hierarchies: if an agent accepts the prospect, it is sent to the member in the next layer (column in the sketch of the graph shown in Fig. 1). If the agent in the last layer also accepts the prospect, it is stored in the final portfolio of this social organization. To model parallel processing, we chose one of the possible ways to extend Sah & Stiglitz' (1986) accept- and reject-rule for polyarchies: if an agent accepts a prospect, it is sent to one in the next layer and possibly stored in the final portfolio. If an agent rejects a prospect, it is sent to the next immediate neighbour. Only if all agents within a layer have all rejected a prospect, it is dumped for good. Note here that the chosen dynamic rule for polyarchies ensures that members belonging to the same hierarchical layer are endowed with equal decision-making competence.

The purpose of the example shown in Fig. 1 above is to illustrate that some social organizations may indeed support the individual's ability to screen information. The perfect screening function is the step function shown in Fig. 1. If agents were equipped with this

perfect screening function, there would be no need for any structure to support their decision. But accepting Simon's assumption of bounded rationality, individuals are at best capable of a limited screening of information. This is modelled by imputing a screening function to the agent that deviates from the step function of perfect screening. The more the deviation, the more severe the condition of bounded rationality. Moreover, a non-monotonic screening function could be used to model "traps." For the purpose of the present example, however, agents are equipped with a monotone screening function as shown in Fig. 1.

The example is so constructed that the graph screening function, representing the social organization's ability to screen information, dominates the individual's agent screening function for any value of prospects different from zero. That is, no matter what the distribution of prospects, as long as all the mass of the prospect distribution does not cumulate at zero, the social organization represented in Fig. 1 would always do better than a comparable collection of independent agents, subject to the constraint that the costs of organization are not excessive.[2]

If we think of the above example in terms of the individual's ability to screen the distribution of fitness de- and increasing prospects offered by the environment within which the agent lives, we have by way of example shown that it is possible, in some instances, that the social organization helps structure the information flows among individuals to increase their individual fitness. In relation to Simon's theory of altruism, the significance of this example lies in the possibility that individuals who organize in a social group may accumulate knowledge, symbolized by the composition of the final portfolio in the above example, that would give each member of the group a fitness advantage compared to a situation where the members are independent individuals. This social organization would

[2] For this example, the agent screening function was defined as $f(x)=(\tanh((x-1)/dx)+1)/2$, $dx=2.068564969066327$. The analytical expression of the graph screening function, a polynomial of degree seventeen, was derived by a recursive procedure provided by Christensen & Knudsen (2002a). In this case the

accumulate knowledge that left docile individuals with a fitness advantage over the individuals that had to instruct themselves or were less susceptible to instruction.

We have thus shown that a positive docility term upon which Simon's theory of altruism pivots would indeed be possible, even in the ancient and prehistoric times long before effective media of storage and transmission of social knowledge had been invented. Whether the social groups actually adopted such fitness enhancing structures is an open empirical question. Since it can be shown that structures of two members (Sah & Stiglitz, 1986; 1988), or any finite number of organization members (Christensen & Knudsen, 2002b), exist in which the screening of the social organization dominates the individual's screening ability in either good or bad times, Simon's theory of altruism appears to rest on very secure grounds.

Discussion

The above review of Simon's selection theory, according to which docility evolves to breed successful altruism, indicated that this theory rests on a very secure ground and therefore deserves serious consideration from biologists as well as social scientists. It was further indicated that Simon's selection theory is remarkable in pointing to a different mechanism to bring about altruism than the kinship selection and structured population models that have emerged from Hamilton's (1964; 1975) work.

As we have seen, Simon's theory is consistent with Hamilton's. According to Simon's model, we should expect altruism to be a pervasive feature of every society, clustered or not. In Simon's model docility gives the individual a possible fitness advantage to be realised in social organizations and societies that transmit useful instructions to individuals. As the fitness advantage of docility increases by society's accumulation of useful empirically tested

---

graph screening function simplifies to $G(f(x))= y^4(y-2)^4(y^2-2\ y+2)^4$, $y= f(x)$. As the reader can verify, according

knowledge, the possibility will increase that a society or a social organization may tax the individual by also conveying instructions that lead to fitness-reducing behaviour.

Integrating Hamilton's (1964) model into Simon's showed that in a clustered society there was room for some level of docility, even if the members of a social organization on average received bad advice. As long as the docile persons would tend to team up with other docile people, and the less docile persons tend to wander off, altruism would evolve, but only slowly. As societies grew increasingly homogenous and accumulated increasingly more sound knowledge, a strong positive selection for docility would eventually occur. Hamilton's (1964) model thus supports Simon's theory of altruism.

When the positive selection of docility picks up speed, Simon's argument, however, relies on the critical assumption that organizations know better than individuals. It is assumed that the individual member of a social organization on average is better off because of the instructions received within the social organization. While this assumption seems self-supporting in the face of the modern scientific and technological advances, it must be remembered that Simon's argument is biological as well as social. Docility is an evolved property of human organisms and other creatures. Therefore, the challenge is to show that a positive docility term would be possible even in the ancient and prehistoric times long before effective media of storage and transmission of social knowledge had been invented. In order to do this, it was argued by example that it is possible, in some instances, that the social organization helps structure the information flows among individuals to improve their individual fitness. As mentioned above, the significance of this example lies in the possibility that individuals organized in a social group may accumulate knowledge that would give each member of the group a fitness advantage, compared to a situation where the members were independent individuals.

---

to this example: $x=0 \Rightarrow G(f(x)) = f(x)$, $x<0 \Rightarrow G(f(x)) < f(x)$, and $x>0 \Rightarrow G(f(x)) > f(x)$.

Simon's model is a biological model, to be sure, but it is defined on the basis of an "even if" argument that also makes it applicable to economics. Simon explicitly relates his model to "neo-Darwinian theory" (e.g. Simon, 1993: p. 156) according to which the criterion for success is biological fitness. Yet Simon (1990: p. 1665) himself regards his argument as an "even if" argument. Even if genes of individual persons are the controlling sites for natural selection, Simon's model identifies a mechanism that selects for altruistic behaviour well beyond altruism to close kin, beyond support from expected reciprocity or social enforcement, and beyond support from positive assortment and viscosity. This mechanism is the transmission of socially learned behaviours and beliefs that can be received by docile individuals. Even if altruistic behaviour is penalized because genes of individual persons are the controlling sites for natural selection, the general capacity to learn at the socio-economic level may contribute with an offsetting positive fitness effect. For this reason Simon's model is applicable to both economics and biology. Note here that Simon's model requires a general capacity to learn at the socio-economic level, but acknowledges additional specific factors that are not present in the model. Such factors include the actual forms of social transmission and learning as well as the actual forms of social organizations and societies that have emerged through history. The emergence of docile individuals may well have favoured the emergence of particular social organisations particularly well suited to benefit from their members ability to receive, accept and believe instructions, which in turn may have helped the further emergence of docility, and so on. This points to an explanatory circle that both involves upwards causation: from docile individuals to new forms of social organisation, and downwards causation: from particular organisational forms to a shift in the expression of docility.[3]

[3] Campbell (1974) and (Sperry, 1969, 1991) provide a description of the concept of downward causation and its use in scientific explanations. See Hodgson (1998) for its potential use in economics.

The possibility that selection may operate simultaneously on docility and the social organization within which the advantage of docility is expressed further indicates an unexplored potential of Simon's theory of altruism. Two promising pathways for further exploration of Simon's altruism theory have been indicated in the course of the above argument. First, Simon's theory deserves to be exposed to empirical test by biologists in a design that allows examination of the veracity of the competing explanations of clustering and docility. Second, Simon's theory deserves to be examined in a design that traces the nested effects of docility and the screening properties of social organizations. Regarding the first point, it is straightforward to base predictions on a classical controlled experiment including the four treatments: a) docile + clustered, b) docile + homogenous, c) less docile + clustered, d) less docile + homogenous. Here "docile" refers to the current characteristic of an organism, e.g. to be inferred by the period it takes a newborn organism to reach maturity. One could therefore select two types of organism and expose these organisms to a "clustered" or "homogenous" environment. If the ranking of the outcome variable altruism were a>b>c>d (or even better, a,b>c,d), we would conclude that Simon's docility effect is probably more important, whereas the ranking a>c>b>d (or better still, a,c>b,d) indicates that Hamilton's viscosity effect is most important.

A number of further implications arise for behavioural economics and theories of economic organization. From the viewpoint of behavioural economics, a particularly important implication of Simon's argument relates to the motivation of employees. Simon repeatedly emphasized that the promotion of organizational identification by instilling organizational pride, loyalty and values has an obvious advantage in motivating the employees to actively pursue the goals of the organization. Because of docility and bounded rationality, Simon's argument established why an employee would not consistently act in economic self-interest but often attempt to aim at realising the goals of the organization.

The implications for current behavioural economics, and in particular the branch known as the new institutional economics, are profound. The new institutional economics relies extensively on opportunism to explain the existence, structure, and boundaries of the firm. Simon's argument, however, implies that altruism as a by-product of docility may significantly benefit firms seeking to exploit the possibility of inculcating the employees with organizational pride, loyalty and values. This argument suggests that the new institutional economics may benefit from integrating the docility assumption along with the assumption of bounded rationality that was always a keystone in its foundations. This indicates that a promising path of development for the new institutional economics would be to consider issues of organizational identification and learning along with issues of contracting.

Finally, and related, Simon's theory of altruism points to the need for further development of the evolutionary theory of the firm, extending some of the shared overlap of Simon (1947), March & Simon (1958), Cyert & March (1963), and Nelson & Winter (1982) by developing a detailed theory of the micro-evolutionary processes that give rise to the evolution of economic organizations. In particular, Simon's emphasis of the relation between docility and organizational identification points to further consideration of organizational routines as the medium in which instructions are stored, and organizational identities as the expression of these instructions. That is, organizational routines, a term introduced by Simon (1947) and further developed by Nelson & Winter (1982), may be viewed as a "replicator," the code containing the accumulated organizational instructions which are replicated when new members of the organization are instructed in organizational procedures, norms and values. And the organizational or social identity may be viewed as an "interactor," the expression of the routines that enables interaction among the members of an organization. This conceptualisation, here invited by Simon's emphasis of docility and his extension of the argument to encompass organizational identification, may be used as the foundation of a

theory of linguistic selection (Knudsen, 2002b) and as an extension of the behavioural aspects of Nelson & Winter's (1982) evolutionary theory.

Conclusion

It is a noteworthy characteristic of Simon's theory of altruism that it is applicable to biology as a complement to the usual biological theory of kinship selection and its recent extensions. Because of the evolved property of docility, the typical human being is endowed with a great potential for being instructed, and because of bounded rationality, the individual cannot discriminate well between the instructions that further selfish goals and the instructions that, at a cost to the individual, contribute to furthering the goals of others. That Simon's theory is biological is noteworthy because it offers a new, convincing and so far under-explored mechanism for the emergence of altruism in biological populations. It is further noteworthy that Simon's model, because it is based on a mechanism of social learning, is also applicable to economics and has immediate implications for the understanding of human nature that invites revision and development of behavioural economics and some of its most important branches, including the new institutional economics and the evolutionary theory of the firm. For these reasons Simon's selection theory deserves further attention. Simon's selection theory is finally noteworthy because it is evidence of his wide scientific reach and taste in scientific problems.

References

Axelrod, R. (1997). *The complexity of cooperation. agent-based models of competition and collaboration*. Princeton, NJ: Princeton University Press.

Bergstrom, T.C. (2002). The algebra of assortative encounters and the evolution of cooperation, *International Game Theory Review*, *Special Issue* (Amir, R., Bergin, J. & Knudsen, T. eds.), Forthcoming.

Christensen, M. & Knudsen, T. (2002a). *The architecture of economic systems: towards a general framework.* Odense, Denmark: University of Southern Denmark (unpublished mimeo).

Christensen, M. & Knudsen, T. (2002b). *Extremal screening, cost and performance of economic architectures*. Odense, Denmark: University of Southern Denmark (unpublished mimeo).

Cyert, R.M. & March, J.G. (1963). *A behavioral theory of the firm*. Cambridge, MA: Blackwell Publishers.

Hamilton, W.D. (1964). The genetic evolution of social behaviour. *Journal of Theoretical Biology*, *7*, 1-52.

Hamilton, W.D. (1975). Innate social aptitudes of man: an approach from evolutionary genetics. In Fox, R. (Ed.), *Biosocial Anthropology* (pp. 133-155). New York: John Wiley and Sons.

Hodgson, G. M. (1998). The approach of institutional economics. *Journal of Economic Literature, 36*(1), March, 166-92.

Knudsen, T. (2002a). The evolution of cooperation in structured populations. *Constitutional Political Economy, 13*(2), 129-148.

Knudsen, T. (2002b). The significance of tacit knowledge in the evolution of human language. *Selection*, Forthcoming.

March, J.G. & Simon, H.A. (1958). *Organizations*. New York: John Wiley.

Myerson, R.B., Pollock, G.B. & Swinkels, J.M. (1991). Viscous population equilibria. *Games and Economic Behavior, 3*, 101-109.

Nelson R.R. & Winter, S.G. (1982). *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.

Price, G.R. (1970). Selection and covariance. *Nature, 227*, 520-521.

Price, G.R. (1972). Extension of covariance selection mathematics. *Annals of Human Genetics, 35*, 485-490.

Sah, R.K. & Stiglitz, J.E. (1986). The architecture of economic systems: hierarchies and polyarchies. *American Economic Review, 76*(4), September, 716-727.

Sah, R.K. & Stiglitz, J.E. (1988). Comittees, hierarchies and polyarchies, *Economic Journal, 98*, 451-470.

Sella, G. & Lachmann, M. (2000). On the dynamic persistence of cooperation: how lower individual fitness induces higher survivability. *Journal of Theoretical Biology, 206*, 465-485.

Simon (1947) [1997]. *Administrative behavior* 4th edition, The Free Press: New York.

Simon, H.A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99-118.

Simon, H.A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63* March, 129-138.

Simon, H.A. (1979). Rational decision making in business organizations. *The American Economic Review*, September, 493-513.

Simon, H.A. (1983). *Reason in human affairs.* Stanford: Stanford University Press.

Simon, H.A. (1987). Rationality in psychology and economics. In Hogarth, R.M. & Reder, M.W., [eds. ?], *Rational Choice: The Contrast Between Economics and Psychology* (pp. 25-40). Chicago: University of Chicago Press.

Simon, H.A. (1990). A mechanism for social selection and successful altruism. *Science, 250,* December, 1665-8.

Simon, H.A. (1991). Organizations and markets. *Journal of Economic Perspectives, 5*(2) Spring, 25-44.

Simon, H.A. (1993). Altruism and economics. *The American Economic Review, 83*(2) May, 156-161.

Simon, H.A. (1997). *Models of bounded rationality* (vol. 3): *Empirically grounded economic reason.* Cambridge, MA., The MIT Press.

Figures

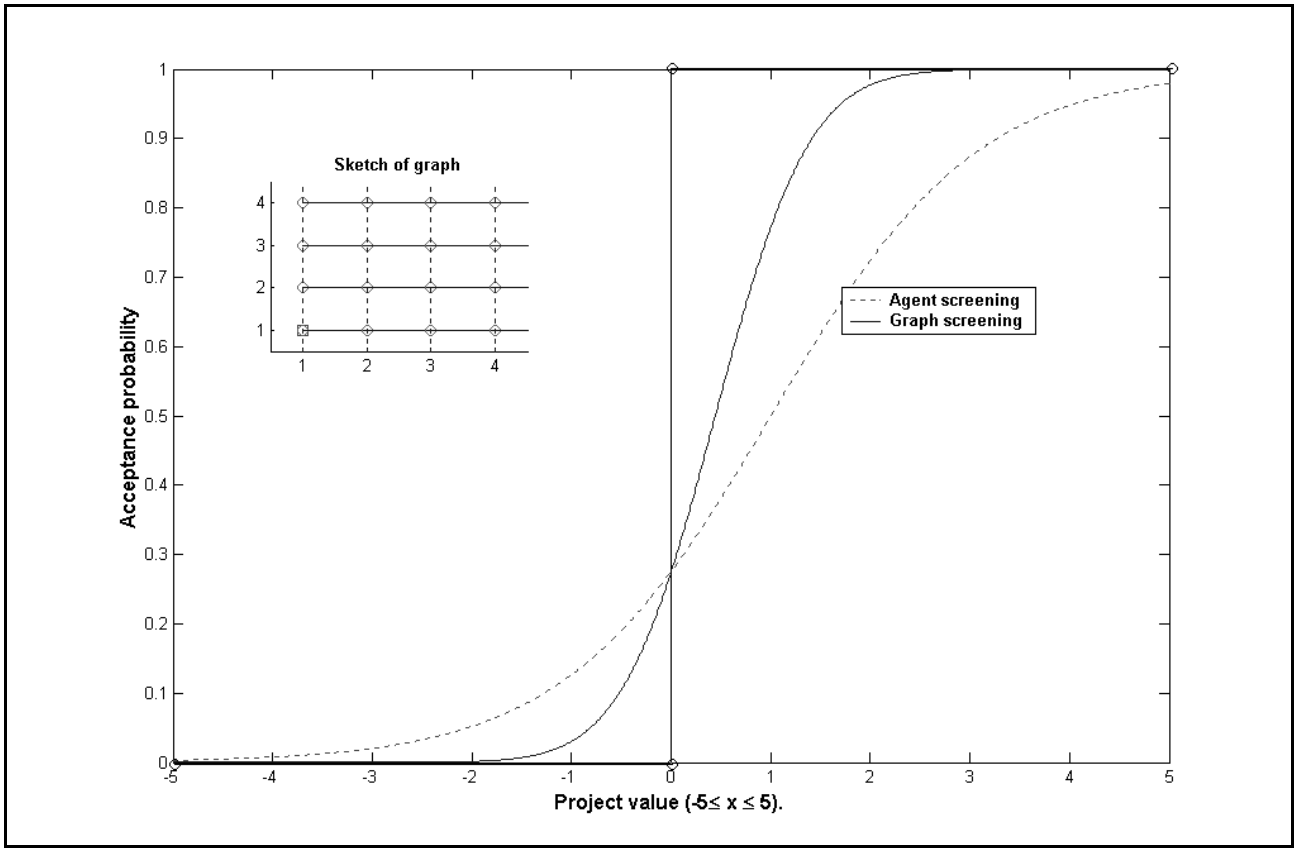

Fig. 1: Agent- versus graph screening. The example was constructed so the graph, for any project distribution, provides better screening than the individual agent, unless all mass of the project distribution cumulates at 0 (in this case the graph and the agent screen equally good).