Economics and Philosophy, **22** (2006) 441–468 Copyright © Cambridge University Press

REVIEWS

doi:10.1017/S026626710621112X

The Stag Hunt and the Evolution of Social Structure, Brian Skyrms. Cambridge University Press, 2004, 149 pages.

In this short but substantial book, Brian Skyrms draws together a number of results from several previously published papers on game theory, signalling, and the origin of social structure, to provide an insightful analysis of the problem of cooperation and collective action. Although this book may be viewed as a sequel to Skyrms' 1996 *Evolution of the Social Contract* – the text on the rear jacket suggests as much – *The Stag Hunt* fails to be a sequel on two grounds. First, a typical sequel extends the story of an earlier work by reworking themes within the same narrow framework of previously established constraints.[1] Second, thanks to Hollywood, we also tend to expect each sequel, upon completion, to leave increasing amounts of uninteresting ground for future exploration. Neither of these statements, in this case, is true. *The Stag Hunt* is much more than a mere sequel.

The Stag Hunt agrees with the general spirit of its predecessor in that traditional problems of the social contract (such as the emergence of fairness and cooperative behavior, and the persistence of both in the face of deviants, to name a few) are better addressed from an evolutionary point of view than from that of traditional decision and game theory. However, whereas *Evolution of the Social Contract* employed the replicator dynamics as its primary tool, *The Stag Hunt* shifts attention away from the replicator dynamics to other models of greater realism and conceptual interest. Two alternative approaches to modelling cultural evolution – local interaction models (the subject of part I) and dynamic social networks (the subject of part III) – dominate the discussion, with the replicator dynamics relegated only to the middle chapters (part II) involving signalling.

The move away from the replicator dynamics serves to address one important, general problem with the methodology of *Evolution of the Social*

[1] *The Godfather: Part II* being the notable exception to the rule.

Contract, first noted for the particular case of the evolution of fair division by D'Arms (1996) and elaborated on by D'Arms *et al.* (1998). If it is the case that "the key to the evolution of cooperation, collective action, and social structure is correlation" (Skyrms 2004: xii), wherefrom does the correlation between behaviours or strategies originate? In *Evolution of the Social Contract*, the author's previous book, correlation between strategies was loaded into the replicator dynamics as a free parameter (see Skyrms 1996: 19). The absence of a plausible, independent story underwriting the introduction of such correlation rendered the explanatory story suspect. What prevented one from invoking correlation of strategies whenever it proved expedient to do so?

In *The Stag Hunt*, the discussion of local interaction models in Part I illustrates how correlation between strategies can naturally arise if interactions between individuals are constrained according to a previously given network of social relations. To block objections that this simply pushes the explanatory story back a step, part III provides a model of how such a social structure may come about from purely random interactions between individuals who modify their behavior according to plausible (and experimentally corroborated) learning rules, with one caveat that I shall return to later. The capstone of part III combines the evolution of social structure with the evolution of strategies, arguing that "the most empirically realistic version of reinforcement learning, fluid-interaction structure and slow imitation decisively and unambiguously tip the scales in favour of cooperation" (Skyrms 2004:122). Very nice.

However, those acquainted with yet other criticisms of *Evolution of the Social Contract* may find their concerns with the overarching project unaddressed. Consider, for example, Kitcher's objection regarding evolutionary game theory's ability to explain morality:[2]

> [I]t's important to demonstrate that the forms of behaviour that accord with our sense of justice and morality can originate and be maintained under natural selection. Yet we should also be aware that the demonstration doesn't necessarily account for the superstructure of concepts and principles in terms of which we appraise those forms of behaviour. (Kitcher 1999)

The Stag Hunt does relatively little to address these charges of overreaching behaviorism. A primary analytic concern remains identifying the basins of attraction for strategies which are behaviorally equivalent to cooperative, or just, actions. When the underlying models have several free parameters, Skyrms identifies – in many cases – how the basins of attraction change as the parameters are varied. Yet the focus on forms of behavior, rather than concepts or principles or sentiments, still dominates.

[2] D'Arms (2000) raises a similar point.

One should keep in mind, though, that the rhetoric has changed between *Evolution of the Social Contract* and *The Stag Hunt* to mitigate the criticism of excessive behaviorism. Whereas the former, for example, claimed to have provided a possible beginning "of an explanation of the origin of our concept of justice" (Skyrms 1996: 21), *The Stag Hunt's* express aim is more muted. The "fundamental question of the social contract," writes Skyrms, is "[h]ow can you get from the noncooperative hare hunting equilibrium to the cooperative stag hunting equilibrium?" If *this* is the fundamental question of the social contract, and there is room here for debate, we *do* receive an outline of a general answer to this question. I suspect, though, that moral and political philosophers will still be generally unsatisfied with the answer, and may want more. I understand entirely; I still want a pony.

CHAPTER 1, AND PART I: LOCATION

One important contribution *The Stag Hunt* makes to the social contract literature[3] is its refocusing of the discussion upon the Stag Hunt, rather than the Prisoner's Dilemma. The Stag Hunt takes its name from Rousseau's story in *A Discourse on Inequality* (I shan't describe the game here) and provides a better reformulation of the problem of the social contract for one crucial reason: although the Prisoner's Dilemma illustrates the conflict between individual optimality and collective optimality, All Cooperate is not an equilibrium. Since social contracts are – at least apparently – stable, it seems odd to begin with a game theoretic model that makes a stable social contract rationally impossible. The Stag Hunt better models the problem of the social contract because the "all cooperate" outcome (All Hunt Stag) is an equilibrium as well as the "all defect" outcome (All Hunt Hare). It does not assume the social contract to be rationally impossible at the outset.

This is why Skyrms claims that the problem of social contract formation becomes a problem about equilibrium selection or, alternatively, equilibrium transition. If a society is in a nonequilibrium state (or, heaven forbid, the All Hunt Hare equilibrium, the Stag Hunt representation of the state of nature), how can that society settle upon a social contract – i.e. how can it move into the All Hunt Stag equilibrium? Here we find another argument against using the replicator dynamics to model social evolution: according to those dynamics, the social contract cannot form through gradual means. If society happens to be in the state of nature (All Hunt Hare), under the replicator dynamics any minor deviation (say, any mildly revisionary proposal to move society more closely towards a social contract) will ultimately be discarded, and society will return to

[3] Although it is not the first, as Skyrms notes.

the state of nature. It doesn't make sense, argues Skyrms, to model social evolution using dynamics which rule out, *a priori*, the gradual formation of the social contract. As such, he suggests, we ought to consider alternative dynamics.

The first alternative dynamic considered, and the focus of Part I, are local interaction models. Suppose people are positioned on the squares of a checkerboard and play a game by interacting with their eight nearest neighbors. People receive a total payoff equaling the sum of the payoffs from each pairwise interaction. After everyone has interacted, players engage in a round of strategic learning and modify their strategies. According to the learning rule known as *Imitate-the-best*, players compare how well they did to their nearest neighbors, and adopt the strategy which received the highest payoff, provided that the highest payoff exceeds the player's current payoff.[4] In contrast, the learning rule known as *Best Response* assumes that one's neighbors will continue to follow the same strategies in the next generation that they follow in the current generation, and selects the strategy which provides the highest expected payoff. Should a tie between several strategies occur, it is broken with a coin flip.

Given this brief sketch, and without discussing the core results, it should be clear how *The Stag Hunt* leaves a great deal of interesting ground unexplored. For example, the local interaction models considered in chapters 2 and 3 involve only square lattices and rings (at the very end of chapter 3, Skyrms briefly considers one alternate structure). One may wonder, do the convergence results depend upon the topology of the local interaction model, or the size of the population?[5] What happens if more sophisticated learning rules are employed? What happens if more than one type of learning rule is present in the population? And so the questions multiply. Yet, to my mind, this is a good thing: in its discussion of local interaction models, *The Stag Hunt* has identified a fertile area of research, and much work remains to be done.

PART II: SIGNALS

Although a natural unity links Part I and Part III, the middle section of *The Stag Hunt* concentrates on the evolution of inference (chapter 4) and cheap talk (chapter 5), and hence functions more as a conceptual detour than a bridge. The discussion of the evolution of inference revisits Lewis' well-known signaling game from *Convention*, and will be familiar to readers of *Evolution of the Social Contract* from the treatment of the

[4] *Imitate-the-best* is a form of dissatisfaction-drivin learning.

[5] Yes, they do. Unfortunately, this margin is too narrow to contain a demonstration of this fact.

evolution of meaning. The chapter on cheap talk illustrates how, somewhat counter-intuitively, meaningless babble can radically transform the basins of attraction of equilibria in games like the Stag Hunt, in addition to creating entirely new equilibria.

While it is undoubtedly important to show how our inferential capacity could have been produced by evolution as a solution to adaptive problems, the chapter on the evolution of inference is the least developed part of the book and does not actually provide a model of how this might happen. Here we find a conceptual sketch of how one might go about constructing such a model. It would be interesting to see this done, primarily to determine whether the envisioned path from proto-truth functions to full-fledged inference is as easily traveled as intimated.

By contrast, the treatment of Stag Hunt and bargaining games with cheap talk is much more thoroughly fleshed out. Cheap talk – the exchange of costless, meaningless signals over the course of play – has been thought to be relatively ineffective in influencing equilibrium outcomes, even by game theorists such as Robert Aumann. To the extent that cheap talk has been considered effective, it has been thought to primarily destabilize equilibria. Skyrms gives the example of a "secret handshake" used to identify cooperators in the Prisoner's Dilemma. If talk is cheap, mutants which extend the secret handshake yet defect can invade the population.

However, Skyrms shows that both of these views underestimate the power of cheap talk. Cheap talk can create new equilibria, and also can change the size of the basins of attraction in the Stag Hunt, such that All Hunt Stag, rather than All Hunt Hare, becomes the state with the largest basin of attraction. Although this latter claim may seem unintuitive, it makes sense upon reflection. It's wrong to think that, with respect to sending signals, the only two possibilities are that the signals be meaningful (in the sense of a Lewis signaling system) or meaningless. If signals are correlated with hunting stag or hare, and strategies allow one to condition one's response upon the signal received, then the *only* time when receipt of a signal fails to convey information is in the special case where all combinations of signals and responses occur with equal frequency. In all other cases, signals – even if they are "meaningless" in the sense of a Lewis-style signaling system – become correlated with response-types, and convey information. The full effects of cheap talk, though, remain an open question.

PART III: ASSOCATION

Part III addresses the second conjunct of the title – the evolution of social structure. This work, carried out in collaboration with Robin Pemantle, considers how social structures, like those appealed to in the local interaction models of Part I, come about. The model is as follows:

given a population, each person assigns a numeric weight to every other player in the population. These weights, if all non-negative (and with a sum strictly greater than zero), can be converted into interaction probabilities by straightforward normalization; if the weights are negative, interaction probabilities can be obtained with a minor amount of fiddling. People choose to play a game with others according to these interaction probabilities. After each interaction, some of the players (either just the initiator, or both, depending on the model) receives a payoff. These payoffs are used by the players to adjust the weights they assign to each other. As the weights increase, and weight tends to concentrate on some players rather than others, interaction structures emerge. The final chapter considers what happens when strategic dynamics (i.e. learning by *Best Response*, *Imitate-the-best*, or other learning rules) are combined with structural dynamics.

Much can be said about the technical content of the final two chapters, which cover a great deal of original and interesting material. Since, however, this would be of little benefit to readers unfamiliar with the book, I shall concentrate on a few overarching philosophical questions which, to my mind, remain unanswered.

First, regarding the caveat mentioned earlier: although we do get, in part I, a story about how social structure influences the evolution of justice and cooperation, and we do get, in part III, a story of how social structure might emerge, how well do the two stories line up? For starters, the interaction structures considered in part I are highly regular: two-dimensional square lattices (or toruses) or one-dimensional lines (or rings). The social structures which are shown to evolve in part III lack any such topological regularity. The emergence of justice and cooperation according to the local interaction models of part I is unlikely, then, if the model of social structure *formation* is that of part III. Moreover, the interaction model of part I is strictly deterministic, with the interactions being hard-wired and always taking place in each generation. How might a probabilistic model of the evolution of social structure like that of Part III be modified to give rise to deterministic interaction structures within a reasonable (i.e. human) time-frame, rather than in the infinite limit?[6]

Second, a different concern regarding the type of explanation offered. The penultimate section of chapter 3 ends with the statement, "[t]he *structure* of local interaction makes a difference" (Skyrms 2004: 42). This sentiment ramifies over the course of the book. We find that *signals* make a difference. The *co-evolution of strategy and structure* makes a difference. The

[6] Notice that the model of social network formation given in chapter 6 might not even converge to a deterministic interaction structure in the limit. In cases where stars form (Skyrms 2004: 90), the interaction probabilities can converge in the limit to a person splitting his time between two individuals.

relative rate of the co-evolution of strategy and structure makes a difference. How are we to assimilate all of these things-which-make-a-difference into a coherent explanatory account of cooperation, collective action, and the problem of the social contract? This remains unclear.

Consider, by way of comparison, the following crude way of putting the meta-narrative of *Evolution of the Social Contract.* The replicator dynamics are a model of cultural evolution. Strategies (i.e. behaviors) which have very large basins of attraction are highly likely to evolve, if the initial conditions were selected at random. Many behaviors of interest, such as fair division in resource allocation problems, retribution in the ultimatum game, and coordination upon meaningful signaling systems, turn out to evolve Lo and behold! often with surprising frequency – under the replicator dynamics. And, thus, we have an explanation for how certain core features of society could have come about through the operation of a blind, dumb, gradual process of social evolution.

It's not at all clear that *The Stag Hunt* permits anything like a similarly crude meta-narrative to be formulated. And perhaps this is a good thing. (Although I'm no postmodernist, I'm suspicious of grand meta-narratives, myself.) However, this also means that it's difficult to see exactly how the explanation works. If everything makes a difference, then all of the various parameters which factor into the evolution need to be set *just so* in order for us to find the behavior we find in society. Yet if only a very small set of possible paths leads from the primordial state to our current social state, what have we discovered? We *already know* that at least one such path exists, namely, whatever path we followed to get us here. If too many things make a difference, the project, then, risks becoming a game-theoretic archaeology for the human sciences (no – not in the sense of Foucault!) instead of a descriptive or predictive science facilitating substantive claims about future social states. At least that's a worry.

No work of philosophy shorter than the completed autobiography of Tristram Shandy can hope to answer all questions, and good work in philosophy raises more questions than it answers. As a snapshot of the recent state-of-the-art in evolutionary game theory, and the philosophical applications thereof, *The Stag Hunt* manages to surpass its predecessor in both scope and content. If only George Lucas had been so successful.

J. McKenzie Alexander
London School of Economics

REFERENCES

D'Arms, J. 1996. Sex, fairness, and the theory of games. *Journal of Philosophy* 93(12): 615–27.

D'Arms, J. 2000. When evolutionary game theory explains morality, what does it explain? *Journal of Consciousness Studies* 7: 296–9.

D'Arms, J, Robert Batterman, and Krzyzstof Górny. 1998. Game theoretic explanations and the evolution of justice. *Philosophy of Science* 65: 76–102.
Kitcher, P. 1999. Games social animals play: commentary on Brian Skyrms' *Evolution of the social contract*. *Philosophy and Phenomenological Research* 59(1): 221–8.
Skyrms, B. 1996. *Evolution of the social contract.* Cambridge University Press.
Skyrms, B. 2004. *The stag hunt and the evolution of social structure.* Cambridge University Press.

doi:10.1017/S0266267106221126

Cognitive Economics. An Interdisciplinary Approach, Paul Bourgine and Jean-Pierre Nadal, eds. Springer, 2004, xiv + 479 pages.

'Cognitive Economics' is a newcomer to economic research. As of now, only a few publications bear its name, and it saw its first European and its first international conference in 2004 and 2005 respectively. This book, carrying the new subdiscipline's name as its title, collects 27 articles from fields as diverse as economics, artificial intelligence, logic, psychology and physics. With many of the articles being surveys, the book serves both as an introduction to the field – nine essays explicitly cover the 'disciplinary bases for cognitive economics' – as well as a 'tool for future research'. The anthology fulfils these two purposes well. Researchers interested in bounded procedural rationality, social influence on individual decision-making and the dynamics of adaptive social systems can learn modelling techniques from outside their fields, and they are offered a wealth of suggestions on how to apply them fruitfully to economic problems.

Cognitive Economics, the editors of this book suggest, is a unified research program that brings a *cognitive turn* to economics. 'It aims to take into account the cognitive processes of individuals in economic theory, both on the level of the agent and on the level of their dynamic interactions and the resulting collective phenomena' (Bourgine and Nadal, v). Now, mainstream economics also takes into account agents' cognition, attributing preferences and beliefs, and modelling deliberation under uncertainty or incomplete information. However, these models rely on two strong assumptions. First, agents are assumed to be *substantively rational*: they deliberate in whatever necessary way to arrive at an optimal choice. Second, the equilibria are presumed to emerge directly from the agents' reasoning. No concrete equilibration process leading to the coordination state is modelled. Rejecting these assumptions, Cognitive Economics focuses on the agents' cognitive constraints, and the deliberative procedures resulting from these limitations. Because agents' information processing capacities are limited, substantive rationality in the sense of universal optimising capacities is excruciatingly costly or simply unreachable. Instead, agents employ *cognitive procedures* that yield at least good-enough results for specific environments. Which specific procedures the agent employs will strongly influence her behaviour in changing environments; hence the procedures themselves and the way agents acquire them through different kinds of learning will be of interest to economists.

Of course, cognitive science has pursued this type of research for more than four decades. Economists have ready access to these results, and do not need to replicate the research. But cognitive science is often criticised for failing to model human thought as inherently social, and here Cognitive Economics makes its central contribution. Boundedly rational agents in general cannot coordinate their actions with the actions of others in such a way that the optimal equilibrium is instantly reached. Recent research, instead, has investigated possible *trajectories* towards equilibrium in these cases – agents *learn* to adjust their behaviour in repeated interaction in order to achieve a social optimum. But these learning models are based on individual rationality: each agent has the optimum as her goal, and searches (with limited capacities) for information so as to adjust her behaviour to best reach this goal (Kalai and Lehrer 1993; as they point out, their model presents learning not as a goal in itself, but as an implication of utility maximisation). Cognitive Economics, in contrast, develops models where agents strictly rely on certain deliberation and learning rules, without having the overall optimum as a goal in mind, but where aggregate behaviour still converges towards this optimum. Cognition thus is not solely cognition of the individual. In fact, 'individual behaviour…is not [Cognitive Economics'] main subject of interest' (Walliser, 196). Instead, Cognitive Economics often models cognition as *distributed*: as information processing distributed over a large number of individuals, who interact in social networks, and influence each other. Consequently, it studies economies as *complex adaptive systems*, and investigates their stability conditions, adaptation dynamics and equilibrium paths.

The book is divided into three parts: three programmatic essays, nine introductory essays and 15 essays on areas of advanced research. The economic introductory essays (essays 2-4) mainly review textbook material, and focus surprisingly little on issues of interest here, be it non-expected-utility, epistemic justifications of game equilibria or models of learning. Makinson's essay on non-monotonic reasoning (essay 6) discusses qualitative logics that allow inferring more conclusions from a set of premises than is classically authorised. His essay provides a wealth of structure that modellers of individual reasoning may find very useful, and it also clarifies the relation of these structures with classical logic. Unfortunately, no other essay in the book makes use of these results (a fate that this essay shares with essay 13 by the same author on conditional statements and directives). In particular, it is a pity that the relation of qualitative logics to logics of belief change remains unexplored, despite their importance for decision and game theory.

Alexandre and Frezza-Buet (essay 7) give an overview of several classes of numerical AI models used for modelling human cognitive abilities. Genetic Algorithms (GA) support determining optimal cognitive procedures for a well-defined search space by simulating an evolutionary process (these may be familiar from evolutionary game theory). Artificial Neural Networks (ANN) are often used to model associative learning. They consist of functioning rules that define the computation performed by the network's units, learning rules that specify how the units' and network's parameters are adapted as a result of learning, and the architecture that defines the way units are connected. Cases of ANN where agents are modelled as units and learning as social influence will be discussed below. Stochastic Behavioural Models, and more specifically Markovian Decision Processes (MDP), allow modelling complex reinforcement learning, where reward is delayed. All these models are part of the numerical paradigm. They stand in contrast to symbolic models, exemplified in standard decision theory, belief revision, indeed the whole propositional attitude tradition, which uses systems of symbol manipulation to model cognitive processes. The authors contend that numerical techniques are 'better adapted to such fields as economics, where expertise and knowledge are too often unconscious and hard to formalise precisely' (Alexandre and Frezza-Buet, 114). Unfortunately, they do not provide more arguments for this interesting but controversial claim.

The research topics section covers individual deliberation, market dynamics and social networks. Starting with individual deliberation, Orléan (essay 12) proposes a concept of collective belief that focuses on the group as an autonomous entity. To believe that group *G* believes *p*, according to this proposal, means to believe that the majority of *G*'s members believe that group *G* believes *p*. As the author notes, such a concept of collective belief has been fruitfully applied to pure coordination games. Here, each agent faces the problem of identifying the *salient* equilibrium out of many equilibria that every player can identify as such. According to the proposed concept of collective belief, an agent *A* chooses equilibrium *E* because *A* believes that all players believe that the group believes in the salience of *E* – not because *A* believes that all players believe in the salience of *E* on the basis of their beliefs in the others' individual tastes and believes. Choosing on the basis of this kind of collective belief, *A* has an advantage. It is far more plausible that cultural traits and group identities are common knowledge (think of stereotypes) than that individual tastes and beliefs are. Hence, players choosing on the basis of their beliefs about what the group believes will lead more readily to a coordinated result than players choosing on their beliefs about what all the other players' individual beliefs are. Orléan then proposes to transfer the concept of collective belief to investment decisions in financial markets. In financial bubbles, he claims, a 'strange and enigmatic' disconnection between individual and collective beliefs occurs. Investors may individually believe that an asset is overvalued, but continue to buy, because they believe that the market will continue to rise. Employing the new concept of collective belief, Orléan suggests, helps to explain the bubble without having to assume the presence of irrational agents.

However, the avoidance of irrationality comes at a price. Financial markets, after all, are not games of pure coordination. In coordination games, collective beliefs are stable because they are self-enforcing. Once agents come upon a collective belief that supports a coordinated result, they have no reason to deviate from such successful collective beliefs. Such a self-reinforcement does not exist in financial markets – bubbles eventually burst, destabilising any collective belief that led to their existence. To invoke collective beliefs in explaining financial market dynamics means invoking common knowledge of *p* until *p* isn't common knowledge anymore. Such an explanatory strategy remains *ad hoc* until an explicit model of the dynamics of such a collective belief is provided.

Tallon and Vergnaud (essay 14) develop a non-standard expected utility (EU) model that does not require the sure-thing principle but satisfies the requirement that the decision maker positively values information. They follow Hammond's strategy of justifying the axioms of their EU model by deducing them from axioms of dynamic choice, but relax consequentialism and instead derive their model from *separability* and *selection of optimal strategies* (SOS). Whereas consequentialism requires that an agent's choices are identical in a decision tree and its strategic equivalent form, SOS only requires that the choice in the decision tree is a subset of the optimal strategies in the equivalent strategic form. The 'weak sure thing principle', derived this way, only requires that if an agent prefers betting on *AÈC* to betting on *BÈC*, and both *A* and *B* are disjoint with *C*, then the agent also prefers betting on *AÈD* to betting on *BÈD* for all other events *D* that are disjoint with both *A* and *B*. The authors further show that with their axioms, the agent always has a positive value of information – a conclusion that distinguishes their model from other non-standard expected utility models. It is difficult, however, to see the relevance of this weakened sure thing principle. The authors point to a family of models that employ possibility measures for a qualitative description of uncertainty, which violate the standard sure thing principle while satisfying its weak form. The real point of contention, though, which set off the whole debate about the principle in the first place – Ellsberg's Paradox – remains a powerful counterexample to the weakened sure thing principle.

Turning to markets, Kirman (essay 18) argues that rational collective behaviour often cannot be directly related to rational behaviour of the collective's individual members. At the example of the Marseille fish market, he shows how the aggregate demand exhibits a standard downward sloping relationship between prices and quantities of fish transacted, while the demand curves of the studied individual buyers does not exhibit this standard property. Thus, the aggregate data can be rationalised under the standard rationality axioms, while individuals constituting the aggregate do not behave rationally. To explain the macrophenomenon, Kirman concludes, one cannot employ a 'blown up version' of the microbehaviours. Instead, he suggests an analogy between human institutions and a beehive or an ant's nest. There, individual ants' or bees' cognitive abilities are strictly limited. They operate in a restricted neighbourhood, obtaining most of their information from those with whom they interact. Despite the simplicity of their behaviour and their reasoning rules, however, the aggregate outcome of their behaviours is surprisingly sophisticated. Following the analogy, Kirman suggests modelling the economy as a complex system, drawing on techniques from statistical mechanics.

Thankfully, the book provides comprehensible introductions (essays 8 and 9) to the most important class of these, the Ising model. Originally developed for the explanation of ferro-magnetism, this model allows inferring interesting properties of a system that cannot be deduced from the bare properties of the system's components. This invites its application to social phenomena, as long as the system in question can be described by something structurally equivalent to the energy function. Phan *et al.* (essay 20) survey and extend some microeconomic models that use the Ising model to investigate *social influence* on individual decisions. The standard model of an agent's discrete choice is turned into an Artificial Neural Network (ANN) by including a social influence component as an additive element to the private utility component. One such model they discuss takes the agents' utility as the equivalent to the energy function of the original Ising model, and models social influence as the agent's adaptive expectations of her neighbours' choices. All agents i simultaneously maximise V with their binary choice (buy, $\omega_i = 1$, not buy $\omega_i = 0$):

$$V_i = \max \omega_i \left(h_i + J_\vartheta \sum_{k \in \vartheta} \omega_k - p \right)$$

where h_i represents the individual preference of the agent, J_ϑ is the social influence factor from the agents in the neighbourhood ϑ, and p is the price of one unit. That is, their individual choice makes V_i positive if the agent buys and null otherwise. This model shows interesting 'avalanche' effects. Take for example an incremental price decrease. This will make some additional agents buy – namely those whose individual evaluation $h_i - p$ has changed given the price change. But those who choose to buy for that reason will change the situation in their neighbourhood ϑ, changing the social influence on some of the neighbours. Some of those influenced will consecutively also buy, further changing the social influence and possibly triggering significant change in the whole population. The model's capacity to capture such processes are of great interest; however, the results that the authors present require quite restrictive assumptions on the models, e.g. symmetry of social influence, same size of neighbourhoods, and very specific distributions of preferences.

The models discussed assume that all agents are connected to a local neighbourhood of homogenous size. In his survey of Agent-based Computational Economics (ACE), Phan (essay 22) shows how this restricting assumption can be relaxed by introducing *social networks*. A social network represents the interconnectedness of a population of agents, specifying each agent's 'neighbour' with the help of a graph. A network is called *regular*, if each agent is connected to his closest neighbours. Through increased random replacement of connections, a network loses its regularity in degrees. Neighbourhoods of complex social systems can be modelled as such social networks. Their stability against external shocks or entropic disturbance, and their dynamic behaviour out of equilibrium, depends on the network's degree of regularity. Through simulations, ACE explores these connections between network regularity, the system's stability conditions and its out-of-equilibrium dynamics.

Zimmermann (essay 23) discusses a further expansion of these systems. He envisages a notion of social learning that goes beyond the ability of agents to adjust in the light of their neighbours' influence. It allows the agent to revise the existence and strength of her neighbourhood links as a consequence of the evolving degree of affinity she feels for, or credibility she accords to, her different neighbours. Agents reallocate their 'closeness' to those neighbours who most frequently have agreed with them in the past. Simulation of an evolving network starting from randomly drawn closeness connections after 10.000 steps of learning then yield an interesting result: a very small number of agents have the power to trigger large avalanches at the level of the whole population. 'Expert leaders' have emerged due solely to their structural position, as the result of a social process.

This ambitious book contains many more interesting chapters on viability theory, stochastic game theory, the evolutionary analysis of communication, social influence in social choice, strategic models of coalition and network formations, a dynamic voter behaviour model in a population with bimodal conflicting interests and a discussion of cognitive efficiency of social networks, which unfortunately cannot be covered here. Instead, a few critical remarks are in order.

The programmatic essays make promises that the later chapters do not bear out. For one, they claim that Cognitive Economics studies *both* adaptation *and* reasoning processes implemented by economic agents in their interactions. The claim that 'it conserves both the time of evolution and the time of eduction' (Bourgine 2) suggests a synthesis of the two approaches. But the book offers no such synthesis; rather, it puts numerical and symbolic models of cognitive processes side by side, and the many chapters treating complex interactive systems focus on numerical models alone. Of course, it may well be that numerical models also are capable of modelling reasoning, but little discussion beyond a brief proclamation of Alexandre and Frezza-Buet can be found. In fact, one author warns that 'cognitive economics, which provides powerful models separately in an eductive and an evolutionist perspective, fails at this time to provide an integrated analytic framework of reference' (Phan 393).

Further, the introductory essays claim that Cognitive Economics can be empirically justified: 'cognitive economics is not armchair economics. The links between cognition, evolution and institutions must be tested by means of field surveys, laboratory experiments, computer simulation and the analysis of models' (eds., vi). Consequently, the book offers two chapters on experimental studies. Politzer (essay 5) surveys the relatively well-known shortcomings in individual reasoning and decision-making, and he cautions about the methodological soundness of many experiments of this sort. The results he surveys are clearly an inspiration for cognitive economists, but do not show how the cognitive economists' own models can be put to the test. Noussair and Ruffieux (essay 19) survey experimental research on markets. Again, the behavioural patterns they report invite the construction of new models; but they say little about how such models of complex interactive systems can be put to the test. In the concluding essay, Lesourne addresses this issue: 'Having to describe complex stochastic processes, the model builders are compelled to introduce numerous assumptions concerning the sequence of events, the way in which information is drawn, the data concept in memory, the size of adaptations, etc.... There is a risk of multiplying ad hoc models based arbitrarily on debatable assumptions' (Lesourne, 468). It is helpful to contrast Cognitive Economics with Behavioural Economics in this regard. Both programmes acknowledge that economics rests on *some* sort of implicit psychology; both relax simplifying assumptions for greater psychological realism and modify other assumptions to acknowledge human limits on computational power, willpower and self-interest. Behavioural economics, however, constructs models that are 'generalizations of standard ones' (Camerer and Loewenstein 2003: 47). Cognitive Economics, as the book portraits it, is quite willing to accept substantial deviation from the standard models. Further, Behavioural Economics justifies introducing psychological assumption as an improvement of economics *on its own terms*: 'The ultimate test of theory is the accuracy of its predictions' (Camerer and Loewenstein 2003: 5). Cognitive Economics concentrates more on *plausible* models of cognition, at the price of less empirical testability and verification – as expressed in one of the programmatic essays: 'the aim is not so much to explain certain realized phenomena as to show that certain phenomena are possible' (Walliser, 196). This need not be a disadvantage, and indeed may be a necessity in order to develop the new programme – but it should be made clear that the models are not well empirically founded at the current stage.

Lastly, many of the essays would have greatly benefited from a scrupulous proof reading. Repeatedly, sentences are fragmented, graph and section references are mismatched, and bibliographic references are incomplete. This can make for a frustrating read.

All in all, this anthology gives insight into a fascinating research area. It confronts a cognitive science approach with the explanatory aims of the social sciences, and for this purpose presents interesting novel modelling techniques. It will be of interest to a wide range of researchers from cognitive science, economics, the social sciences and AI.

Till Grüne-Yanoff
Royal Institute of Technology, Stockholm

REFERENCES

Kalai, Ehud and Ehud Lehrer. 1993. Rational learning leads to Nash Equilibrium. *Econometrica* 61(5): 1019–45.

Camerer, Colin F. and George Loewenstein. 2003. Behavioral economics: past, present, future. In *Advances in behavioral economics*, ed. Colin F. Camerer, George Loewenstein, and Matthew Rabin. Princeton University Press.

doi:10.1017/S0266267106231122

Just Work, Russell Muirhead. Harvard University Press, 2004, 209 pages.

As its title suggests, *Just Work* is an attempt to articulate an account of the justice of work. But the title is not just descriptive, it is also imperative – just work! – and this leads Muirhead to examine the question, why we work. Muirhead argues that although Americans work for the instrumental reasons of monetary and material sustenance this cannot entirely explain our working life.[1] For besides how much money we make, we also evaluate our work in terms of how it fits us – the way it brings meaning to our lives through developing our talents and capacities. For Muirhead, the justice of work requires not just that we fulfill socially useful roles, for example teaching as opposed to stealing, but also that our work personally fits us in a certain way.

[1] This book is specifically directed to Americans and American working culture, but it draws on sufficiently broad concepts of liberal democracies in general to be accessible and interesting to a wider audience.

Marx is famous for bringing to our attention the expressive and formative experience of our working life, as well as the alienation that occurs when our work is at odds with our distinctly human capacities. He understood the source of alienation in terms of capitalist modes of production and he sought a solution to the problem in a social revolution, which would return production to the hands of the workers. The apparent failure of Marxist-type experiments in the modern world has meant that many reformers and politicians forego issues of fulfillment in work and focus instead on contractual relations between employer and employee. Contractual relations are attractive because they require mutual consent, specifically an employees consent, and thus express a respect for individuals as autonomous human beings, while at the same time they fall short of the coercion and bias involved in naming individuals to jobs to which they are 'ideally' suited. Liberal democrats do not deny that our jobs are formative, but they leave it up to the individual to decide which jobs to take and when to leave them; that is they leave it up to the individual to decide who they want to be.

Thus Muirhead's ideal of a personal fit with one's job is not new, but as an element of the justice of work in liberal democracies, it is contentious. In the first few chapters of *Just Work* Muirhead argues for the relevance of personal fit to liberal democratic notions of justice and in latter chapters he explores what such a personal fit might involve. In what follows I want to canvass some of his reasons for situating personal fit into the philosophy of liberal democratic justice, briefly look at a few different interpretations of personal fit, and finally introduce Muirhead's solution.

Muirhead argues that at the heart of liberal democracies exists a tension between, on the one hand, the affirmation of the individual and her ability to conceive of and follow her own conception of the good life, and on the other hand, the fact that just democracies cannot be neutral "between ways of life that contribute to economic productivity and those that do not." (Gutmann and Thompson 1996: 280) Just democracies ask that we are self-reliant and that we do our share of the work to keep ourselves and society afloat. Whatever this amounts to, just democracies cannot support a class of people who exist because of the work of another class. But the affirmation of work is potentially at odds with the affirmation that we are free to conceive and follow our own conception of the good life. Because jobs undeniably shape the kind of people we are and the sort of life we live, and because not all jobs are open to all people – due to, for instance, education, natural ability and luck – the working life represents a constraint on our ability to conceive and follow a conception of the good life.

That our freedom as democratic citizens is constrained in various ways is not controversial, but just how and when those freedoms are constrained is. Muirhead's contribution is to suggest that our freedom to conceive and follow our own conception of the good life may be reconciled with work if our work fits us; if we could endorse our work as part of the good life. This does not require, Muirhead argues, a Marxist restructuring of the economy or governmental control of what counts as the good life – in other words, the ideal of fitting work need not overshadow the freedom it was meant to save.

Muirhead further argues that if we ignore the ideal of fitting work, and focus – as most liberals do – on informed consent, freedom of exit and mutual benefit, then we cannot articulate the real problems with many jobs. Muirhead offers up a myriad of convincing examples to make this point, but he tends to focus on the problems of domestic service, be it servants of the late nineteenth century or the nannies of today. For instance, think of the woman who leaves her own family in the Philippines to travel to the UK to care for another's family. She comes willingly, receives a wage which she hopes to send home so that her children can get an education, and so she and her husband can one day afford a home. Moreover she can quit whenever she likes, she receives time off from her work and the family in the UK treat her well. What is the problem with such work?

The problem for Muirhead seems to be that while in one sense, such a person can be said to have chosen this job, her choice is predicated on the fact that her other options were even worse – perhaps the inability for her family to stay together, the inability to pay for basic education and healthcare etc. Thus even though she made a choice, it is not a choice that represents her conception of the good life. Just because she chose the best of a bad lot does not mean that she deserves what she chose.

Liberal democracies recognize that we each have a claim on a life that is, in some sense, our own. But in order to realize that claim we need more than the formal freedom to choose, we also need to pay attention to the choices available to us. Choice is important not for its own sake, but because it makes it more likely that we will find fitting work. But this crucially depends on there being valuable things to choose from. To realize our own conception of the good life we need a set of options that allow for the cultivation of our human capacities and individual purposes. Just what those capacities and purposes *are* is a matter for debate, but some will come from the conception of a person implicit in liberal democratic theory. For instance, jobs need to support and not undermine a sense of dignity and equality something that, arguably, is lacking in the above example.

If liberal democracies can support some notion of personal fit or meaningful work it must be one that allows for individual difference in what one might find meaningful and one that is potentially open for all to achieve. In the last four chapters of his book Muirhead looks at different ways the concept of fitting work has been understood. He begins with the Protestant work ethic and the notion of a calling. Here one finds fitting and meaningful work by exercising the talents and capacities given by God and intended by his design to serve the greater good. You are called to your work by God and it is a sacred duty to perform it faithfully. All honest work, however common, is filled with divine purpose. You work, not for riches and comfort, but to fulfill an obligation to God, and in the face of predestination, to manifest signs of salvation. But as America became more secular this meaningfulness of work was lost. While in a God-fearing world we work to fulfill God's purposes now we just work from habit and to fulfill the desires that money can meet.

From the Protestant work ethic Muirhead turns to Mill's philosophy for a secular articulation of what might count as fulfillment or meaning. Muirhead seems to appreciate Mill's emphasis on the development of higher human capacities, and the importance of individual choice to their development. But for Mill fulfillment of those capacities only comes to those with the courage, foresight, dedication, education and talent to disregard social expectation and live by the light of their own choices. Because these qualities are rare, those individuals capable of fulfillment are also rare. Muirhead concludes that this conception of fulfillment is too limited for a democratic culture, which understands meaningful existence as possible for all citizens.

From Mill Muirhead turns to Betty Friedan's careerism. Echoing Muirhead's earlier concerns with domestic labor, Friedan's solution to the lack of meaning in such work is to turn housewives into career women. Friedan imagines a working world in which personal purposes find articulation in the purposes of one's work. This ideal fit between an individual and what she does as completely expressive of who she is reminds us of Marx. Although Freidan does not call for a restructuring of the economy, for Muirhead, the problem with both of them is the same: in requiring work to be the sole source of expressive meaning they set the bar so high that no job can attain it – whatever the economy. Even the best careers are still jobs and require discipline as well as creative expression.

When we make the notion of fit unrealistic, for instance, because it requires a metaphysical belief that few people still hold, or because it entails a life very few people can acquire, or because it asks that all careers fit us completely, then the category of fit appears impossible and the tendency is to disregard it for something less difficult to implement. In light of this, Muirhead suggests that the most realistic way to think about fitting or meaningful work is in terms of Alasdair MacIntyre's understanding of a practice. For an activity to be a practice it must be coherent, complex, cooperative and socially established, but most importantly it must motivate the people who do it in terms of goods that are internal to the practice. An internal good is something that is acquired through learning to do an activity well, and which only through dedication to the activity can one grasp the goods involved.

An internal good is never fully understood outside of participation in the activity in which it is made manifest, but this is not to say we cannot recognize internal goods that we personally do not have. For instance, we can recognize the particular grace and prowess that comes from being an athlete, or the sense of style and design that comes from working with art. But the point is that internal goods are the sort of things that, through dedication and practice, shape us in their image, and then insofar as we identify with them, they become expressive of who we are. This is different from a model of fit or meaningful work in which we first have a passion or talent for something, and only then look for a job to engage it. On the model of internal goods the job, via its internal goods, creates the passion or particular talent.

A good fit requires that a job have internal goods to acquire, that we indeed acquire them and finally that the goods we acquire shape us in ways that coincide with our understanding of the good life. Unlike the Protestant work ethic, Mill's fulfillment or Freidan's careerism, work as practice connects the ideal of fit with the actual experience of work. You do not need to be particularly smart, wealthy or courageous to achieve the internal goods specific to a job – practicing law as well as fixing cars and teaching philosophy can bring internal goods to the right person. You do not need to believe in God to find purpose in your work, the purpose now comes from how the work allows you to be a certain kind of person, to express yourself in ways that you identify as good. And finally, work as practice does not promise a perfectly expressive experience, rather it says that there is hope to partially enrich our lives with our work. But that enrichment will take discipline and hard work, and even then it is only work – one aspect of our life, not the entire thing.

If the justice of work is to include the concept of personal fit or meaningful work, then it is essential that we have some way of identifying work that fits our human capacities and our democratic policy. It is only if we can identify such work that we can begin to present all individuals with choices of value and thus give ourselves and others the opportunity to honestly pursue respective conceptions of the good life. But the problem with this is twofold. First, as I mentioned earlier, the internal goods which develop our capacities are never fully understood from outside a practice and although we can sometimes recognize them we cannot be sure to always appreciate them. As Muirhead cautions we must be careful not to dismiss or attempt to reform a job merely because we are unfamiliar with its internal goods. To combat this tendency Muirhead says that we must sometimes accept on faith the testimonials of those who do the job. Beyond just accepting what others say, he suggests that we might look at the habits that the job instills and try to evaluate how they contribute to doing the job well, but also how they contribute to living well in general. This brings us to the second difficulty: a liberal democracy requires that different conceptions of the good life coexist, but if we evaluate a job's internal goods partially in terms of the life it enables, then we are always in danger of dismissing a job because it enables a life which we do not understand or do not appreciate.

In the case of work as practice it happens that on both levels, that of evaluating the job itself for internal goods and how various internal goods instill qualities for general living, we encounter places where personal, cultural or societal prejudices may make seeing the "good" of the internal good difficult. I do not think Muirhead would deny this although he says little about how we might deal with it. Nor do I think it is a problem that necessarily undermines his suggestion that we think of meaningful work as a practice. But certainly it is a difficultly that deserves more attention and one that we might have expected Muirhead to address in a more straightforward fashion especially since this is the crux of the problem with any attempt to implement substantive accounts of the good into theory. One solution for these sorts of difficulties is to engage in hermeneutic dialogue, but it is unclear what particular form it might take here.

Overall, *Just Work* is an interesting and ambitious attempt to grapple with the myriad concerns that affect us in our working life. Muirhead is to be credited for his unflinching insistence that personal fit is necessary to the justice of work and indeed necessary to liberal democracies and their conception of a person. This is especially true in the current climate where Americans vote in ever greater numbers for representatives who privilege market forces over government controls in the workplace. But just for this reason we may wonder if Muirhead's arguments about the value of choice in relation to freedom can really overcome, what seems to be the dominant belief, that the market itself can make us richer and more free. Although I agree with so much of what Muirhead says, I worry that it is not enough to alter our present course.

Leah M. McClimans
London School of Economics

REFERENCE

Gutmann, A. & Thompson, D. F. 1996. *Democracy and disagreement*. Belknap Press.

doi:10.1017/S0266267106241129

Classical Utilitarianism From Hume to Mill, Frederick Rosen. Routledge, 2003, xiii + 289 pages.

With this book, Frederick Rosen, for many years the director of the Bentham Project at University College London, presents a resolute defense of classical utilitarianism. It proceeds, in the first part, by tracing the development of three fundamental concepts – utility, justice, and liberty – in the writings of Hume, Helvétius, Smith, Bentham, Paley, and John Stuart Mill, looking at all of them from the perspective of the Epicurean tradition. The second part consists of four shorter, less historical and more philosophical, chapters, three of which address some of the most prevalent criticisms of utilitarianism – that it provides a justification for punishing the innocent, for sacrificing the individual in order to promote the general happiness, that utilitarianism cannot give an adequate foundation for minority rights – and one chapter on Isaiah Berlin's claim that Hobbes and Bentham shared the same concept of negative liberty. The 11 chapters of the first part make up four fifths of the book, which leaves one with the impression that the second part is of relatively less importance, especially since all four chapters had previously been published in the form of articles. As the author notes, the book provides "some of the ingredients" (p. x) for a defense of utilitarianism, rather than representing a systematic defense itself. The book does not, in other words, present a vision of classical utilitarianism, or even a comprehensive survey of the development of the doctrine, but rather a series of shorter presentations and arguments, primarily intended to clarify misunderstandings of classical utilitarians, the role of utility in Adam Smith's writings, and, in various contexts, of Bentham's utilitarianism. Rosen points out that some chapters will be of particular interest to some readers: "historians of economic thought will find most to interest them in the chapters on Smith and Bentham; moral philosophers might find most in the two chapters on Mill, while intellectual historians might be most interested in the material on Hume and Smith" (p. 3). Being most attracted by the moral philosophy, I will concentrate on Rosen's reading of Mill.

Rosen focuses his discussion of Mill's utilitarianism on the relation between Mill and Bentham, and the role of Carlyle. I am not aware of any better discussion of, especially, the relevance of Carlyle to Mill's writings, and greatly enjoyed reading Rosen's arguments.

It is almost a commonplace in the literature, both among contemporary utilitarians as well as their opponents, that Mill rejected Benthamite utilitarianism, which he had seen to be deficient during his mental crisis (around 1827), and presented an alternative conception about 35 years later, in *Utilitarianism* (1861). Rosen claims that even if Mill might have dismissed many of Bentham's views in his essays of the 1830s, later in life he had returned to Bentham's theory. In fact, it is not so much that Mill accepted Carlyle's criticisms – they first met in 1830 – but that by the time he wrote *Utilitarianism*, he felt the need to defend *Benthamite* utilitarianism against these objections. Rosen points out, interestingly, that Mill himself might have given his readers the wrong idea about his intentions when claiming that Bentham had said somewhere that "*quantity* of pleasure being equal, push-pin is as good as poetry" – even though Bentham had never said that. But, possibly, it is only because Mill says that Bentham had used this expression (even if only in his less-well-known 1838 essay on Bentham), that, when seeing Mill's talk of the "quality" of pleasures, he is read as explicitly making reference to this particular passage in Bentham.

Be that as it may, as Mill's discussion of "quantity and quality" are notoriously difficult to make sense of, Rosen would have to present both a new interpretation of Mill and show how this coheres with Bentham's utilitarianism for his claim that Mill defended Bentham to be fully substantiated. In my opinion, Rosen fails to give such a comprehensive reading of Mill. There is no doubt that Bentham was not a "shallow quantitative" utilitarian, yet, to argue for this, as Rosen successfully does, still leaves the problem of conceptually integrating their views: what do "quality" and "quantity" in Mill stand for? Though Rosen illustrates Bentham's uses of these words, he fails to give a clear account of how the relevant passages in Mill are supposed to be read as a result. Rosen does mention that Mill emphasizes the pleasures of the intellect, and convincingly shows how earlier Epicureans, including Bentham, had done the same. It seems to me that at least among any persons remotely sympathetic to utilitarianism, this part of the debate has already been settled. Indeed, so as to make sense of Mill's apparently problematic assertion that some pleasures are better due to their intrinsic nature – which Rosen reads as equivalent to them being *intrinsically* better – Rosen seems to suggest that even for Bentham, the different "qualitative" dimensions of the assessment of pleasures (e.g. intensity, duration, certainty, etc.) cannot straightforwardly be integrated into one summary assessment, even if any two pleasures might be comparable as to their scores on any of these dimensions. Even though Rosen does not present it thus, this seems a very radical reinterpretation of Bentham. In my opinion, Rosen doesn't do enough to show how this suggestion can be squared with Bentham's other writings. Ultimately, then, the project of showing how Mill has returned to Benthamite utilitarianism relies on a very unusual reading of Bentham, and thus only relocates the problem from interpreting Mill to interpreting Bentham.

I want to end with a note of unreserved praise. Rosen suggests that Mill's reply to Carlyle focuses on one particular criticism: the Puritan idea that men could (and should!) do without happiness altogether, and, hence, that the empirical theory at the base of utilitarianism – that people desire only happiness intrinsically – is false. Mill accepts the empirical observation, that people are not motivated by the prospect of their own pleasure, but defends utilitarianism by claiming that the martyr, who sacrifices his life, does so justifiedly only if having the intention of promoting the happiness of others, thus, though apparently not intrinsically or exclusively desiring his own happiness, still performing an action worth of the utilitarian's approval. I think this is an interesting reading of Mill, and as far as I am aware Rosen is the first to claim that *this* part of Mill's utilitarianism is explicitly addressed to Carlyle. A further step would be to examine whether this is ultimately compatible with the psychological theory it is supposedly based on.

There is a lot of material in this book that anyone interested in utilitarianism would benefit from mulling over. It presents interesting perspectives on some traditional themes, and succeeds in clarifying, if not resolving, a number of issues that I am sure will still be discussed in many years.

Christoph Schmidt-Petri
Witten/Herdecke University and University of Glasgow

doi:10.1017/S0266267106251125

Distributional Justice, Theory and Measurement, Hilde Bojer. Routledge, 2003, xv + 151 pages.

Vilfredo Pareto deserves the credit for the first thorough statistical analysis of the distribution of income (Persky 182). Based on taxation statistics from England, Italian cities, several German states, Paris and Peru he claimed to have discovered a law of the distribution of income that revealed the same, unequal distribution of income in political units of very different size, institutional set up,and history. Based on this he claimed that the distribution of income depends on human nature rather than economic organisation and political institutions, and armed with his law claimed the futility of socialist and other egalitarian attempts to change the income distribution. Pareto's claims provoked a host of empirical rebuttals, criticism of his method and conclusions, and alternative ways of assessing the distribution of income. In fact, today the measurement of income and its distribution has become an institutionalized exercise in many countries, and impressive attempts have been undertaken to track inequality trends even on a global level (Milanovic). As a result, the measurement of the distribution of income has come to influence political perception, and how we think about the distributive justice of countries or even larger phenomena such as globalization. Yet, already the work of Pareto immediately revealed that the measurement of inequality involves difficult questions regarding the unit of measurement, and the representation and interpretation of measurement. Even if today the measurement of income has the weight of the factual and the support of refined statistical methods, it does not follow that income distribution is of primary concern for understanding inequality and distributive justice. This raises questions of distributive justice, but there are remarkably few books that would deal with both the theory and measurements of (in)equality. Hilde Bojer's *Distributional Justice, Theory and Measurement* deserves credit for discussing these two themes together in one book.

The book is divided into two parts: first theories of justice, then questions of measurement. The first part introduces major theories of distributional justice including utilitarianism, Rawls' theory of justice, Dworkin's approach and the capability view, as well as libertarianism and Marxism (Chapters 2–7). Generally, the author introduces the main features of the respective approaches, followed by a discussion of objections. The exception to this pattern is one very short chapter on Marxism and libertarianism, about which the author openly acknowledges her lack of enthusiasm (p. 3 – as this chapter neither introduces recent developments in either of the two approaches, and as neither of the two approaches play any apparent role for the subsequent discussion of measurements, the chapter can probably be safely skipped). Bojer adds a special chapter on children and their mothers in the first part, which is especially valuable as the approaches mentioned above generally have little to say on this topic (Chapter 8). The second part of the book is devoted to the measurement of inequality, and starts off with a clarification of the notoriously difficult concepts of income and wealth, then turning to a discussion of the temporal dimension of measurement as well as the unit of measurement (both individual versus household, and questions concerning the components of income, Chapters 9–11). She then introduces general features of economic equality measures, discusses specific examples such as Atkinson's and Kolm's inequality measure, the visual representation of inequality with the Lorenz curve as well as the Gini coefficients (Chapters 12–14). The book closes with chapters devoted to the measurement of poverty, and the decomposition of inequality measures. Two appendices deal with uncertainty and expected utility, and sampling errors respectively; a brief section with suggested further readings concludes.

Bojer offers a succinct discussion of, and a rare chance to learn about, both the theory and measurement of distributive justice. Her courage in bringing together these largely independent and vast literatures is admirable, and contributes to a critical understanding of current discussions of inequalities in the world. As already indicated by the structure of the book, Bojer moves from theories of distributional justice to questions of inequality measurement in the second part. However, it is not the case that this second part directly discusses the measurement implications, and implementations, of the theories of justice discussed in the first part. It could therefore be useful to first read the second part, and then read the first part with the knowledge of empirical measurement issues in mind. Bojer states the reason for this break at the beginning of the book, where she observes "*the gap* between what the philosophers write, and what is studied in empirical analyses of income distribution" (p. 1, italics added). She attributes this gap to the welfarist outlook of empirical researchers and the lack of concern of philosophers "with how their concepts can be made operational for empirical analysis" (p. 2): "philosophers tend to be long on profundity and short on empirical measurement" (p. 86). Her book sets out to tackle this, i.e. "to bridge the gap between moral philosophy and empirical research" (p. 2). But her conclusion is just as honest as it is negative: "I have not succeeded; the bridge is still unbuilt" (p. 2).

As this "gap" and the need to bridge it are of major concern for Bojer, let me turn in more detail to this task, which is made difficult not least due to the complicated relation between theory and measurement in political philosophy. What sort of "thing" is this "gap"? Is the relation between distributional theory and measurement primarily one of a movement from the former to the latter? Surprisingly, Bojer offers few explicit reflections regarding her stated objective of bridging the gap, and I will therefore use the remainder of this review to work out some clarifications. To wit, there is no such thing as *the* "gap," there are many gaps, and keeping this in view may help us think about the theory and empirical analysis of distributive justice.

Let me first turn to the "gap" between distributional theories and measurement, choosing as an example what Bojer has to say about Rawls' theory of justice as fairness. Rawls defines primary goods as those which every rational human being wants, and lists basic liberties, freedom of movement and choice of occupation, powers and prerogatives of offices and positions of responsibility, income and wealth, and the social bases of self-respect as primary goods. According to Bojer, "Rawls, like many other philosophers, seriously underestimates the difficulties connected not only with the practical measuring, but also of the theoretical definition of income and wealth" (Bojer, 44). How income and wealth are measured depends not least on the purpose of the measurement, which in this case must "in some sense and to some extent answer the question: is the distribution just" (p. 86)? In Rawls' case this requires knowledge as to whether the distribution of the primary goods of income and wealth is to the benefit of the least advantaged. Moreover this is not an interest in income as the result of choices, but as a constraint and requirement for everyone's equal opportunity to determine their lives. Finally, the constraint on choice at issue is that of the constraints on life-prospects, rather than on individual consumption possibilities in a shorter, e.g. typically annual, period.

Given these specifications, how then does Rawls underestimate the difficulties of theoretical definition and practical measurement? The specifications just given lead one to expect that *individual life-time* prospects have to be measured. But governments in practice measure household income. Not only is household income important for assessing the standard of living of children, but income as a constraint on choice depends not just on one's personal income, but also on that of the other persons in the household. A shared household reduces costs due to "public household goods," and at least to some extent income is likely to be shared between the members of the household. "It is therefore standard procedure in analysis of personal income distribution to analyse the distribution of *household income*" (p. 78, italics added). However, how households pool their incomes differs from household to household, with immediate consequences for the opportunities of its members. There is therefore a tension due to the household measure, which seems at once a more accurate measure of income than individual wage income, and a measure concealing pertinent questions of equal opportunity such as independent earning power and the opportunity of leaving an abusive marriage (p. 88). Another problem regarding income as a primary good is due to the fact that in practice income is not manna from heaven, of which everyone is likely to prefer more rather then less, but rather something that requires work, and therefore implies a reduction in available leisure time. While it is possible to estimate a full income, which takes leisure into account, Bojer seems to be right when she says that full income is hardly an uncontroversial idea (it requires making leisure and work "somehow" commensurate), and is at any rate "not at the moment in practice observable" (p. 45). Considering such criticism, I would argue that the prima facie plausibility of this part of Rawls' primary goods approach depends on the amphiboly of income as a good *and* a measure in his theory of justice, i.e. the plausibility of income as a good owes much to income as a measure, where the latter due to the work of economists and statisticians has become something that we tend to accept as an objective measure and that is already used for the making, the communication and the monitoring of policies. A "gap" emerges because of the scrutiny of critics such as Bojer, who drive a wedge between income as a good and income as measured.

But what about the philosophers following and reacting to Rawls? Do they show insufficient concern for empirical analysis (p. 2)? Consider Amartya Sen's criticism that the primary goods approach overlooks the differences among people that result in people having very different effective choices based on the same resource bundle. Does this criticism show no concern with empirical analysis? As the criticism is pertinent to a basic concern of measurement, i.e. the unit of analysis, for it to count as not engaged in empirical analysis seems to require in the domain of distributive justice the presumption that empirical analysis means (large-scale) statistical measurements, paradigmatically of the income distribution and the various representations this gives rise to (such as the Lorenz curve, the Gini coefficient, and the other inequality measures discussed by Bojer in the second part). If this is assumed, then "mere" criticism of a unit of measurement is not sufficient for empirical analysis, alterative measurements have to be developed. But this presumption also tends to suggest that theories of distributive justice could not stand in a *critical* relation to measurement (as opposed to the implementation relation, which seeks to move from the theory to the design of appropriate measures). In this critical relation, the theory of distributive justice may simply question the moral accuracy of large-scale measurement in general, or even just particular measurements. In the latter sense, Sen can be read as pointing out that the focus on income offers only distorted information regarding equality, which is moreover not just a criticism of Rawls, but by implication also puts a question mark over distributive politics focusing primarily on income transfers as a distributive tool. Indeed, as income measurement is historically due to the state's need for revenue, it would almost be a miracle, if it just so happened that this revenue tool also serves best for achieving a fair distribution.

And yet, such a critical relation to measurement hardly seems sufficient. Is there not a need for alternatives modes of operationalization and empirical analysis? In the light of the critical relation between theorizing about distributive justice and empirical analysis, two points need to be taken into account. The first one is the trivial, but consequential fact that measurement in large political units, traditionally the "nation state" and now increasingly "the world," requires institutions that gather this information, and people who accept to be "measured." Post-revolutionary France offers ample illustrations of the difficulties of introducing liberal democratic measures, where influential parts of the population still stick to traditional, hierarchical categories and therefore refuse to be counted as equals (Bourguet). If the operationalization of a theory for measurement in the political domain depends to some extent on the social world (or needs to change the social world), then the building of "bridges" is also to some extent the building of a "social world." There is therefore a potentially problematic limitation involved, if the problem of the "gap" is conceptualized as merely one of developing appropriate "measurements." Second, this becomes even more apparent, once it is clearly stated that there is no one gap between theory and measurement, and that in particular not all theories of justice may require large-scale measurements. Bojer quotes Sen's statement that the capability approach requires "a generally accepted list of valuable functionings (that) must be determined by open, public debate and reflection" (Bojer 49). But the demand for open debate coupled with a sensitivity for differences between people and their circumstances puts a question mark over the need for, and possibility of, large-scale measurements that would allow one to identify the least advantaged by means of the yardstick of income. Rather, "open and public debate" seems to call for the creation of relatively local processes for an accurate understanding of equality (and as such processes hardly exist today, it so to speak requires a novel "social world"). Put differently, empirical analysis will in this case be much concerned with questions of democratic process.

As a result, it seems to me that Hilde Bojer puts the spotlight on an important question regarding the relation between theory and empirical analysis in the field of distributive justice, and that she offers valuable discussions of the works of philosophers and economists. To think further about the objective of bridging the gaps between distributive justice and empirical analysis, it may be valuable to take into account the epistemic background out of which the discussions of distributive justice in the second half of the twentieth century have emerged so as to uncover the assumptions producing the "gap," and in particular so as to give the call for measurement one place among other questions of political scale, democratic institutions and process that would correspond to a wider notion of empirical analysis required for the task.

Rafael Ziegler
McGill University

REFERENCES

Bourguet, M.-N. 1987. Décrire, compter, calculer: the debate over statistics during the Napoleonic period. *The probabilistic revolution. Volume 1: Ideas in history*. Eds. Krüger, Lorenz, Daston, Lorraine J., Heidelberger, Michael. MIT Press. 305–16

Milanovic, B. 2005. *Worlds apart. Measuring international and global inequality.* New Jersey: Princeton University Press.

Persky, J. 1992. Retrospectives. Pareto's Law. *Journal of Economic Perspectives* 6 (2):181–92