

JAAKKO KUORIKOSKI

EXPLAINING WITH EQUILIBRIA⁶⁸

Equilibrium explanations are pervasive in economics and biology. They are also becoming more frequent in the social sciences due to the widespread adoption of economic (and sometimes biological) explanatory models. For example, economist Edward Lazear has recently (2000) claimed that the adherence to equilibrium concepts is a major factor contributing to the ‘scientific’ status of economics. It therefore explains to some extent why economic explanatory models have invaded other fields of social inquiry. This view is echoed amongst others in some branches of political science, where microeconomic explanations based on equilibrium solution concepts are seen as the paradigm of scientific thinking, and the demonstrable lack of stable equilibria in models of voting systems are viewed as a devaluation of the field to the status of dismal science (e.g., Riker 1980; Austen-Smith and Banks 1998). However, equilibrium explanations have so far attracted surprisingly little philosophical attention. An interesting exception is the puzzling role of the so-called general equilibrium models in economics.⁶⁹ However, these intricate mathematical constructs are widely held to provide more of a consistency proof to the usual neo-classical modeling assumptions than actual explanations of events or phenomena (Hausman 1992; Weintraub 1985). Moreover, when equilibria are discussed, the actual philosophy of explanation tends to get buried under the mathematics, and therefore a more approachable account might prove useful.

The focal point of this discussion is the account of equilibrium explanations given by Elliott Sober. Sober claimed in his ‘Equilibrium Explanation’ (1983) that equilibrium explanations constitute a counter-example to ‘the causal thesis of explanation’, because they do not require the exposition of the actual causal history of the event to be explained. The point argued in this paper is that explanations of singular events are indeed causal, even those supplied by equilibrium models. What Sober construes as explanations of singular events by putative disjunctive causes should be seen instead as constitutive or structural explanations of (macro) properties of the system (population) under scrutiny. However, equilibrium *models* can be used to explain different kinds of explananda. What *can* be explained with a given explanatory model will in turn be assessed by using James Woodward’s theory of explanation (Woodward 2000; 2003). It is further argued that these various explanatory uses of equilibrium models require that the equilibrium models pick out

⁶⁸ I would like to thank the Finnish Cultural Foundation for its generous support of this research.

⁶⁹ However, the opinion is that the whole theory of general equilibrium is dead (Ackerman 2002).

causally relevant properties of the system and that the models can and should therefore be seen as special cases of causal models. This argument supports Brian Skyrms' contention, namely that it is the dynamics of the equilibrium model that bear most of the explanatory burden (Skyrms 1997).

1. EQUILIBRIA

Intuitively, the idea of an equilibrium is quite simple: it is a stable state of a system that is being maintained as a result of interacting forces in the system. Usually, equilibria are also thought of as having domains of attraction larger than the state itself, so that if perturbed, the system would right itself and return to the equilibrium state. However, as we will see, technically this is not a necessary property of an equilibrium. Why are such states explanatory? Sober's account of equilibrium explanation goes as follows: if a system has an equilibrium with a large enough domain of attraction, the *occurrence* of an equilibrium state can be explained by exhibiting it as an equilibrium. This description is explanatory *because* it shows that many different initial conditions could have led to the state in question. According to Sober, equilibrium explanations thus represent a counter-example to the claim that the explanation of a single event requires the citing of its actual causes, since equilibrium explanations seem to make the actual causal history of an event irrelevant: in other words, the system would have ended up in the equilibrium state even if the causal history had been different in a number of ways (Sober 1983, p. 207).

Sober's example of an equilibrium explanation is Ronald Fisher's argument accounting for a 1:1 sex ratio common in a wide variety of populations. If a population diverges from the 1:1 sex ratio, there will be a reproductive advantage favouring pairs producing the minority sex. The ratio of male-to-female progeny has an impact on a parent's fitness by virtue of the number of grandchildren produced. If males are in the majority, individuals producing female offspring will, on average, have more grandchildren than others. Therefore, any heritable variants that cause overproduction in the minority sex tend to increase in frequency. Thus, the population is eventually driven back to the 1:1 equilibrium (Sober 1983, p. 201-202).

For Sober, the crucial element behind the special explanatory properties of equilibria is the many-to-one relationship between the initial conditions and the explanandum. Therefore, Sober's characterisation of an equilibrium explanation applies to a fairly large number of different models in different domains. When discussing equilibrium models, a distinction should first be drawn between dynamic and non- or quasi-dynamic models. Equilibria in traditional, rational choice -based game theory (usually Nash-equilibria), are usually only quasi-dynamic. An equilibrium is characterised only by its intrinsic properties (as a best-response strategy profile) and little or no attention is given to the question of how the system is supposed to reach it. As Nicola Giocoli puts it (2003), the ascendance of game theory and the associated topological techniques have transformed the paradigm of economic equilibrium from a system of forces to a mutual consistency of abstract

relations (plans and strategies). These non-dynamic equilibria pose additional questions about the explanatory worth of the equilibrium concept, especially if the model has multiple equilibria.⁷⁰ Therefore, I will not discuss these non-dynamic models here.⁷¹

In contrast, the classic Cournot-game of oligopolistic competition, ordinary price-taking partial equilibrium models, equilibria in models of physical systems and models in evolutionary game theory are all dynamic in a loose sense, i.e., they purport to show, or at least to account for, how and why the equilibrium is supposedly reached. Phenotypic optimality models can be used to approximate genetic equilibrium models and therefore can sometimes also be regarded as dynamic equilibrium models when selection is at least implicitly assumed to drive the phenotype into optimum (Esher and Feldman 2001). Evolutionary games, population models and some economic equilibrium models are also dynamic systems in a stricter formal sense, i.e., they are constructed as systems of differential or difference equations. Equilibria are then defined as fixed points, points of the state-space where the rate of change of the system is zero.⁷²

The explanatory worth of equilibria is usually taken to depend on the different stability properties they possess. An important distinction should be made here between two kinds of stability: dynamic and structural. Dynamic stability refers simply to the internal properties of the system as given; is the equilibrium only locally stable or perhaps asymptotically or even globally stable? In other words, given some initial state, where do the given dynamics carry the system? Structural stability refers to the stability of the system itself, i.e., whether qualitative features of the model are robust under changes in parameters or dynamics. These qualitative features usually refer to the topological properties of the model, i.e., a model is structurally stable if minimal perturbations result in topologically equivalent models (Skyrms 1999). There is some controversy about which topological stability features are interesting or even requisites for an acceptable model. Some bifurcation phenomena may be artefacts of the model, and some may reflect physical properties of the modelled system. According to the so-called stability dogma, only structurally stable models can be given an empirical interpretation. According to a more liberal

⁷⁰ Hence, the proliferation of solution concept refinements that attempt to pin down the one ‘right’ solution concept that would always yield a unique solution (cf. Harsanyi and Selten 1988). Bicchieri (1995) argues that the problem of multiple equilibria should be addressed by explicitly modelling the learning processes of the players.

⁷¹ This omission also gives one more reason to evade the thorny question of the explanatory role of general equilibrium theories, because the usual Walrasian *tâtonnement* ‘mechanism’ or its variants cannot reasonably be taken to describe the actual dynamics of actual systems (Rosenberg 1992, Chapter 7). Fixed-point theorems used in the existence proofs of general (in essence game-theoretical) equilibria leave the relevant mapping unspecified.

⁷² However, Skyrms has noted that dynamic equilibria cannot in fact be taken as principal explanatory concepts, because there are many fixed points that are unstable and do not have domains of attraction and therefore do not support the apparently required many-to-one relationship. Skyrms claims that the obvious candidate for an explanatory concept is the attractor (although I would presume attracting sets would be enough). Furthermore, all attractors are not fixed points. (Skyrms 1997) I basically agree with Skyrms, but I continue to speak of equilibria in order to stay true to Sober’s intentions and to the common usage of the word rather than to the terminology of applied mathematics. Besides, the formal finesses of fixed points and attractors do not play any role in the discussion to come.

interpretation, only structurally stable *properties* of dynamic systems should be considered as physically relevant. (Guckenheimer and Holmes 1986, Chapter 5)

2. MODELS AND PHENOMENA

A large part of Sober's paper is devoted to the question of the plausibility of disjunctive token causes; if equilibrium explanations explain events in the same way as ordinary (singular) causal explanations, can alternative histories together constitute a reasonable causal explanans (Sober 1983, p. 204-206)? However, I take this explicit concern with the explanation of individual events to be rather curious. Is the intended target of Fisher's argument really a 1:1 sex ratio of a single population at a given time? The perplexing explanandum is undoubtedly the pervasiveness of the 1:1 sex ratio across different populations, species and environments. What is being explained is primarily a generic pattern or a phenomenon, as discussed by James Bogen and Woodward (1988). According to Bogen and Woodward, phenomena fall into many different ontological categories: particular objects, objects with features, events, processes and states. By explaining phenomena rather than singular events or data, scientists avoid having to tell an enormous number of independent, highly local and idiosyncratic causal stories. From this perspective, the effort put by Sober into discussing the viability of disjunctive causes of a token event seems largely misguided. Instead, according to Bogen and Woodward, phenomena require something they call systematic explanation, which is usually supplied by explanatory models. Fisher's argument is an outline of a model providing a basis for such a systematic explanation (cf. Seger and Stubblefield 2002, p. 7).⁷³

Systematic explanation and singular causal explanations do differ but are by no means independent. Explanatory models as such do not explain individual events, but they can be used to construct such explanations. As Sober himself notes, what *the* cause of a single event is is a context-dependent question, usually dependent on largely pragmatic concerns. Sober also claims that the dilution of the causal requirement to the claim that causal explanations provide 'information about the causal history of the event' amounts to trivialisation of the whole concept of causal explanation (Sober 1983, p. 202-203; see also Gjelsvik in the present volume). Yet, whereas these points lead Sober to sink into pessimism about the whole idea of causal explanation, Woodward's theory of explanation offers a more constructive approach by way of an account of how these context-dependent matters can be met - an account that tells what *can* be explained on the basis of given information or a given model.

The basic idea behind Woodward's theory is that explanatory models show how systems work. By giving a schematic account of the dependency relations in a system or a mechanism, explanatory models codify answers to what-if-things-had-

⁷³ Indeed, in their introductory text, Seger and Stubblefield claim that it is difficult to think of any other biological field concerned with specific phenotypes that would be more model-driven than sex-ratio research.

been-different questions, i.e., assertions about various counterfactual, contrastive states of the system. In the most straightforward cases, these dependencies are represented as functional relationships between variables, where the possible values of the variables form the relevant contrast classes with respect to possible explanatory set-ups. More specifically, models of causal mechanisms give answers to questions concerning the effects of hypothetical interventions on these variables. Woodward's claim is that both systematic explanation and singular causal explanation involve these calculations of the effects of possible interventions in the model, but they differ in what kinds of interventions are relevant—what variables should be held fixed in the evaluation of the relevant counterfactuals. Thus, this account captures the heavy context dependency of singular causal explanations that Sober was worried about without shoving everything of interest under the rug of pragmatics (Woodward 2000, 2003).

Woodward's account also seems, at the outset, to account for *why* equilibrium explanations and their many-to-one dependencies create significant understanding and therefore provide 'deep' explanations. It is because they give answers to a particularly large collection of what-if-things-had-been-different questions, i.e., they have a wide counterfactual 'range'. Even if things (initial conditions) had been different, at least not too drastically different, the end result would have been pretty much the same. However, in principle, this explanatory force does not differ from that of other schematic descriptions of mechanisms. Whereas Sober's intuitions about explanations seem to lean towards epistemic or perhaps modal conceptions in Wesley Salmon's taxonomy (Salmon 1984, Chapter 4), the contrastive-counterfactual perspective advocated here is firmly ontic or realist; understanding is created through the exposition of objective mind-independent dependencies, which facilitate action and inference by supporting counterfactuals about hypothetical manipulations.⁷⁴

3. POSSIBLE EXPLANANDA OF EQUILIBRIUM MODELS

So far so good. According to Alexander Rosenberg, the resilience or persistence of a phenomenon has been associated with explanation since Plato. Rosenberg claims that this fundamental intuition motivates the search for explanation of change in terms of underlying persistence, which itself is seen as intrinsically understandable. Equilibria thus provide a natural stopping point for inquiry, because they minimise the amount of inherently problematic change (Rosenberg 1992, p. 205-206). Rosenberg himself does not advocate this view, but claims that such intuitions have

⁷⁴ Another area where the contrastive-counterfactual framework might prove useful is the controversy around the explanatory relevance of different stability properties of equilibria alluded to above. The adherence to the ontic account demands only that the stabilities and possible bifurcations reflect objective features of the system being modelled. Therefore, stability properties *per se* should not be used as criteria of empirical or explanatory significance. Instead, what the relevant stability properties are depends on the specific problem in hand. The contrasts explained by different manipulations of the variables (initial values) or parameters could and should be analysed on a case-by-case basis. However, these questions are not pursued further here.

been historically influential and that they are one of the main reasons why equilibrium explanations are so popular among different sciences. However, after a more careful examination, this view actually seems to run counter to the account of explanation used here; it is the invariance of the dependencies, not the invariance of the actual behaviour of the system, that is responsible for explanation. In fact, since equilibria with large basins of attraction are highly *independent* of their initial conditions, they actually seem to represent a counter-example to a requirement of *genuine* dependence made by Woodward and Christopher Hitchcock. Woodward and Hitchcock claim that in order to be explanatory, a generalisation (or the corresponding dependency relation) has to be invariant under a *testing* intervention, i.e., there has to be a possible setting of the explanatory variables that would have changed the explanandum state from the actual equilibrium state (Woodward and Hitchcock 2003, p. 17). This requirement is vital, since explanatory dependencies are characterised by counterfactual conditionals, and removing this requirement would allow one to generate apparently explanatory pseudo-dependencies at will. However, in the extreme case of equilibrium models with global dynamic stability, such testing interventions would by definition be non-existent or prove to be so drastic that they would disrupt the whole system, thus making it impossible to speak intelligibly of different values of the same explanandum variable (end-state). It seems that with global stability there could not be any proper testing interventions on the explicit explanatory variables and therefore no explanatory power. This is clearly unacceptable, since such stable equilibria are usually precisely the ones regarded as highly explanatory.

If explanations indeed track dependencies instead of persistence, the interesting explanatory relationship cannot be the one between the initial conditions and the equilibrium state, as might first be surmised, and indeed as seems to have been Sober's view. Instead, what the equilibrium state does depend on are the structural features of the system. *Equilibrium explanations are not causal explanations of events but structural or constitutive explanations of system-level properties.* As an illustration, consider a Lotka-Volterra model of a continuous predator-prey system with strong self-limitation for the prey owing to competition in a finite environment, i.e., in the absence of predators the prey population grows according to a logistic equation:

$$dx/dt = r(1 - x/K)x - bxy$$

and the predators respectively:

$$dy/dt = (-c + dx)y$$

where x and y are the prey and predator densities, b the predation rate coefficient, d the reproduction rate of prey population increase and c the predator mortality rate. When projected onto a predator-prey plane, the trajectories of the system under various different initial values look something like the following:

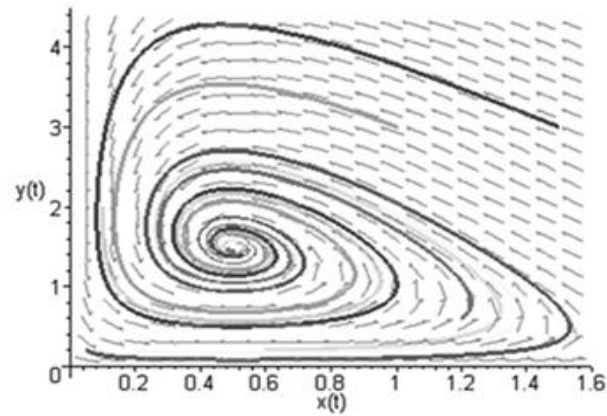


Figure 1. A predator-prey-system with self-limitation

However, if we loosen the self-limitation on the prey, we get the classic Lotka-Volterra model (a is the intrinsic rate of prey population increase), which does not have stable equilibria, but instead yields cyclic behaviour of the predator and prey populations of which there are well-documented instances:

$$\begin{aligned} dx/dt &= (a-by)x \\ dy/dt &= (-c+dx)y \end{aligned}$$

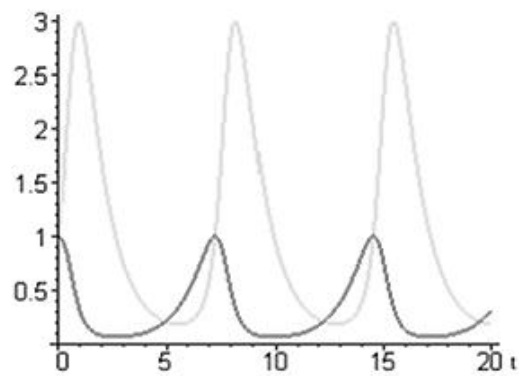


Figure 2. A predator-prey system without self-limitation

Here we have very similar models apparently explaining very different kinds of phenomena. The first is a paradigmatic equilibrium explanation for certain stable levels of predator and prey populations, whereas the second explains cyclic fluctuations of population levels sometimes encountered in predator-prey systems. However, the latter explanation does not exhibit the many-to-one structure of Sober's equilibrium explanation. Moreover, cyclical population patterns extended indefinitely in time are hardly reasonable candidates for token events. Clearly the classic Lotka-Volterra model as such does not primarily explain events. Instead, the explanandum is more naturally interpreted as a property of the system. The predatory-prey dynamics with constants referring to the mean or expected behaviour of token animals (the predation coefficient is sometimes actually meant to be estimable from 'laboratory' experiments with small sample sizes) give a structural (micro) explanation of a behavioural tendency of the macro-system composed of the two populations.

Since it is difficult to see any fundamental differences between the classic model and the self-limiting model yielding an equilibrium, the equilibrium itself is also most naturally seen as a macro-property arising from the constituting causal properties (predation coefficient, etc.) used in the construction of the differential equations. The explanatory dependencies captured by these models are not causal dependencies between successive events, initial conditions and the end state, but constitutive dependencies between properties of a system and their causal basis. Since it would be rather forced to claim that the explanatory principles in these models are radically different (it is also worth noting that the change in the differential equations can be given a loose empirical interpretation), the many-to-one relationship between the initial conditions and the equilibrium in itself is clearly not the factor responsible for the explanatory power in the equilibrium model. Instead, the causal mechanisms and capacities implicit in the model dynamics constitutively explain both the equilibrium and the non-equilibrium explananda.⁷⁵

4. CAUSAL EXPLANATION OF EVENTS WITH EQUILIBRIUM MODELS

Suppose that one insists on demanding an explanation of a token event using Fisher's argument, i.e., an explanation accounting for the occurrence of a 1:1 sex ratio in a population p at time t . The first contrast class of the explanandum event that comes to mind is composed of other (logically) possible sex ratios of p at t . However the initial conditions do not make a difference, either to the sex ratio or to t for that matter. What the argument tells us is that even if there had been some forces acting on the population in the past, forcing the population out of the equilibrium, the population dynamics have *apparently* driven the population back to the 1:1 ratio.

⁷⁵ It should be noted that some regard the L-V models with suspicion precisely because they give rise to qualitatively different phenomena under alterations of parameters and therefore do not satisfy the strict stability dogma (cf. Guckenheimer and Holmes, 1986). However, as has been argued above (footnote 74), the fact that L-V models can account for a variety of qualitatively different phenomena should not be held against them if the changes in the model responsible for the radical alterations can be given an intelligible empirical interpretation.

Notice how the lack of information about the actual history of the event is accompanied by total lack of information that would single out the so-called event of the 1:1 ratio of p at t from any other corresponding pseudo-event (i.e. a 1:1 ratio at t' or t'' , etc.). Contrasting the time index t to other possible values simply does not seem to make much sense. As was argued above, what the argument actually explains is a macro-property of the system, i.e., the tendency of having a 1:1 sex ratio. Once we know this property, explaining unremarkable token events becomes a trivial exercise; the system has a property of being in state s , unless recently perturbed, *and* the system has not recently been perturbed. However, even in this trivial explanation of an event, the rather uneventful causal history of the system makes an indispensable appearance.

Structural or constitutional explanations of properties are not causal in the sense that singular explanations of events are. A property instantiation can of course be given a causal explanation in terms of the causal history of its constituting token structure, but this is an answer to a different question. However, a modified Fisher argument/model could in theory also be used to explain *changes* in the population sex ratio. Fisher's argument relies on an assumption of equal or linear fitness returns from the production of sons and daughters. In fact, as Fisher himself observed, the equilibrium concerns parental effort or 'expenditure', not numbers of offspring. This assumption can be relaxed in a number of ways. The 'costs' of the sexes can differ because of different mortality rates during the parental care period, different resource needs of male and female offspring and different effects on the parents' future reproductive performance. (Seger and Stubblefield 2002, p. 16) Suppose that the environment changes in a way that alters these differential costs. This modification changes the population dynamics and creates a new equilibrium sex ratio (in terms of actual offspring), which the population eventually achieves due to selection. The same kind of exercise could also be done with the Lotka-Volterra model with strong predator self-limitation; a change introduced in the parameters (not in the initial conditions) would shift the equilibrium levels and the populations would eventually be driven to this new equilibrium. In this way similar equilibrium models can be used to explain singular events. In fact, explanations with this kind of structure are the norm in economics and go by the name of comparative statics.

Consider the following simple text-book explanation of a token event using a microeconomic demand-price equilibrium. The explanandum is the increase in the price of steel in the U.S. from 1980 to 1985 by roughly 25% relative to the international (Amsterdam) market level. The explanation for this price increase is taken to be 'voluntary' import quotas imposed by the Reagan administration. In the equilibrium model, the structural effect of the imposed quota on the steel market is represented as a shift in the supply curve (from S_1 to S_2), which moves the market equilibrium along the demand curve D . Market forces are then supposed to drive the price to this new equilibrium level (explanandum), which is higher than the pre-quota level.

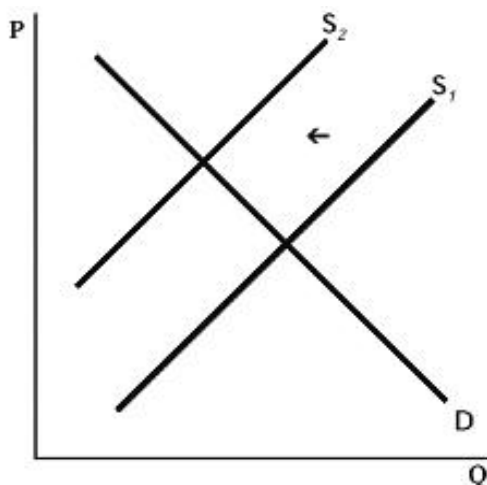


Figure 3. A simple comparative statics exercise

Here an equilibrium model is constructed to represent the state of the system before the event to be explained (sometimes parameters are fitted against data if this is possible), a disturbance is simulated by manipulating the parametrised model and then the explanans is hopefully the (or an) equilibrium of the disturbed model. The disturbance thus causes the event to be explained in a straightforward manner through the mechanism schematised in the model dynamics (in this case implicit and quite informal). More generally, the forces responsible for the attainment of the equilibrium create a counterfactual dependency relationship between structural parameters and the equilibrium state that is invariant under some reasonable interventions on the parameters. Changes in the parameters cause the system to settle into the new equilibrium. This kind of equilibrium *reasoning* underlies the bulk of economic theory used to formulate econometric models usually thought of (implicitly or explicitly) as representing causal relationships. Equilibria here serve more as important characteristics of mechanisms linking variables of interest, i.e., as parts of the explanans, rather than as things to be explained. Mechanisms underwrite the counterfactual dependencies required for causal explanation, but do not as such explain events; they only link them. When explaining token events with equilibrium models, the actual causal history is therefore far from irrelevant.

Daniel Hausman agrees that what the Fisher argument explains, i.e., the sex ratio, is first and foremost a property of the population not reducible to its members. However, he takes this to be a contingent feature of this particular form of argument; if the population were to be divided into two equal subpopulations, one reproducing at sex-ratio of $n:m$ and the other at $m:n$, there would be no selection pressure. Thus, the Fisher argument only yields a population average. Hausman claims that other 'arbitrage arguments', which are essentially equilibrium explanations with the exploitation of disequilibrium (arbitrage) and the resulting selection acting as the

explanatory mechanism, can also be used to explain properties of individuals. Hausman's example of an arbitrage argument is the familiar story used to legitimise the application of rational expectations to the theory of firm behaviour: there are (assumed to be) firms that behave according to predictions made by the best relevant economic theories. Such rational expectations are, on average, assumed to increase 'fitness' relative to firms behaving otherwise. However, there seem to be no systematic differences in firm survival that could be attributed to different predictive abilities. Therefore, on aggregate, firms behave according to the best relevant economic theories (Hausman 1989).

Are properties of individual firms really being explained here? What about token property instantiations or even singular (f)acts? The exact formulation of the conclusion of Hausman's arbitrage argument is as follows: *Ceteris paribus*, the expectations of firms for the most part match the predictions of economic theory. (Hausman 1989, p. 6) More generally, according to Hausman, the conclusion of an arbitrage argument takes the following form: *Ceteris paribus*, in the actual environment almost all x possess H . (Hausman 1989, p. 9) Although ostensibly about members of the population, the conclusion cannot be used to answer questions of the form why *firm x, rather than firm y*, has expectations matching economic theory. In fact, the very impossibility of such attributions served as a premiss in the argument! 'Almost all x possess H ' is just an elliptical way of claiming 'Almost all x that are G (or belong to g) possess H '. What is being explained is still first and foremost a property of the population (g), i.e., a high relative frequency of firms behaving according to economic theory. This conclusion certainly makes it likely that any member of the population will act as predicted, but properties or actions of singular firms would still require some information about their micro-structures or causal histories (cf. Ylikoski 2001, p. 57-60).

5. ARGUMENTS AND MODELS

The preceding discussion was based on a shift of focus from explanations as single arguments to various uses of explanatory models. One motivation for this shift was Sober's perfectly legitimate unease about the multiplicity of causal factors affecting any given event and the subsequent indeterminacy of *the* singular causal explanation of the event in question. In fact, the very idea of contrastive explanandum reveals that something over and above a deductive argument is needed for explanation. This need not mean discarding the deductive ideal or the third dogma of empiricism, at least in one sense: the possibility of deducing the explanandum can serve as a threshold for what can be explained with a given model (Ylikoski 2001, p. 68). What is denied is that explanations can fruitfully be analysed as mere arguments, in isolation from the theories and models used in deriving them.

In addition, focusing on equilibrium explanations as a special and distinct genus of explanatory arguments masks the important fact that the actual fixed points or attractors are not the only interesting features in dynamic systems. If the explanatory force is derived from the many-to-one relationship between the initial conditions and the equilibrium, other interesting and important dependencies in the model do not

get the attention they deserve. For example, George Gilchrist and Joel Kingsolver (2001) have recently stressed that the optimality peaks and the associated population equilibria are not the only interesting features in so-called fitness surfaces and adaptive landscapes. Individual performance or fitness surfaces are mappings from individual phenotypes (or sometimes genotypes) to individual performance or fitness, whereas adaptive landscapes are maps from the mean phenotype or allele frequency to the mean population fitness. These models are necessarily related, but quite distinct, and only adaptive landscapes can be directly related to models of evolutionary dynamics. These models do not merely explain the attainment of equilibrium, but instead can be made to jump through all kinds of explanatory hoops. First of all, the simple fact that selection tends to drive the population up the steepest slope of an asymmetric peak in an adaptive landscape tells us something over and above the equilibrium point itself. However, this assumes the absence of underlying constraints. Asymmetric peaks in a multivariate fitness landscape imply functional, genetic or phenotypic correlations amongst traits that have to be accounted for in the outlining of possible trajectories of the population (Gilchrist and Kingsolver 2001). If the combined fitness surface and adaptive landscape are treated analogously to a description of a mechanism, the explanatory power of these additional constraints can be analysed with the same theory of explanation as that of the equilibrium. Through the various evolutionary mechanisms summarised in the landscapes (selection being only one), these additional constraints can in turn explain various contrasts in the trajectory of the population; different manipulations of these constraints would lead to systematic differences in the trajectory of the population in the adaptive landscape. Moreover, by moving the peak in the adaptive landscape so as to model gradual environmental changes (such as a climate change that translates into a steady change in selection pressure), these combined models can be used to explain directional, long-term changes in populations and thus provide a linkage between link micro- and macro-evolutionary patterns (Arnold, Pfrender and Jones 2002). Gilchrist and Kingsolver aptly start their discussion by imagining a mountaineering guidebook listing the location of every peak and the numbers of climbers encountered at the top. They conclude that without additional information such a book would prove to be of little interest. For a climber, the really useful information would reside in the topography: how steep is the approach, are there ridges leading to the summit, are there other peaks nearby. (Gilchrist and Kingsolver 2001, p. 219)

These considerations show that optimality models and adaptive landscapes can in principle be used to explain at least some general features of actual causal histories of populations. Here the mechanisms underlying the model dynamics are the responsible factor for the attainment of the peak, equilibrium, as well as the way that the peak was actually achieved and the trajectory of the population under possible environmental changes. Therefore, the same theory of explanation could and should be used to analyse these different explananda. Respectively, just because the model providing an explanation is based on equilibrium assumptions does not mean that the explanation itself should have some unique feature distinguishing it from other explanations.

What then is so special about equilibria? If equilibrium explanations - if strictly speaking there are indeed such things - do not themselves possess any special explanatory depth, why are they given such a privileged standing as the hallmark of a truly scientific enterprise? Here is my proposal. Equilibria *are* special and important - just not in quite the way Sober presented. The strength of the concept stems from the ability of equilibrium constructs to underwrite powerful structural dependencies between the parts and the whole *with only minimal assumptions about the constituents*. Consequently, equilibrium assumptions facilitate causal explanations of events by providing *credible and relatively easily tractable mechanisms* linking structural changes and changes in the state of the system under scrutiny. In equilibrium models, there are usually not that many different kinds of nuts and bolts and no intentional effort needed to ensure the obtainment and maintenance of the equilibrium. In contrast, imagine a complex thermostat with elaborate causal feed-back mechanisms maintaining the status quo or a centrally planned economy keeping its books balanced through massive data-gathering operations and active interventions. In a very concrete sense, both of these systems would be in an equilibrium state, but would this equilibrium have any extra explanatory import over and above the elaborate mechanisms maintaining them? Clearly not.

6. REFERENCES

- Ackerman, F. (2002). Still dead after all these years: interpreting the failure of general equilibrium theory. *Journal of Economic Methodology* 9 (2): 119-193.
- Arnold, S.J., M.E. Pfrender and A.G. Jones. (2002). The adaptive landscape as a conceptual bridge between micro- and macroevolution. In A.P. Hendry and M.T. Kinnison (eds.): *Microevolution Rate, Pattern, Process*. Kluwer: 9-32.
- Austen-Smith, D. and Banks, J. (1998). Social Choice Theory, Game Theory and Positive Political Theory. *Annual Review of Political Science* 1: 259-287.
- Bicchieri, C. (1995). The Epistemic Foundations of Nash Equilibrium. In Little, D. (ed.) *On the Reliability of Economic Models*. Kluwer: 91-137.
- Bogen, J. and Woodward, J. (1988). Saving the Phenomena. *The Philosophical Review* Vol. XCVII (3).
- Eshel, I. and Feldman, M. W. (2001). Optimality and Evolutionary Stability under Short-Term and Long-Term Selection. In Orzack and Sober (eds.) (2001): 161-190.
- Gilchrist, G. W. & Kingsolver, J. G. (2001). Is Optimality Over the Hill? The Fitness Landscapes of Idealized Organisms. In Orzack and Sober (eds.) (2001): 219-241.
- Giocoli, N. (2003). *Modeling Rational Agents*. Celtenham UK: Edward Elgar.
- Guckenheimer, J. and Holmes, P. (1986). *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*. New York: Springer.
- Harsanyi, J. C. and Selten, R. (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge MA: The MIT Press.
- Hausman, D. (1989). Arbitrage Arguments. *Erkenntnis* 30: 5-22.
- Lazear, E. (2000). Economic Imperialism. *The Quarterly Journal of Economics* 115 (1): 99-146.
- Orzack, S. H. and Sober, E. (eds.) (2001). *Adaptationism & Optimality*. Cambridge University Press.
- Riker, W. (1980). Implications from the disequilibrium of majority rule for the study of institutions. *American Political Science Review* 74: 432-446.
- Rosenberg, A. (1992). *Economics – Mathematical Politics or The Science of Diminishing Returns*. The University of Chicago Press.
- Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton NJ: Princeton University Press

- Seger, J. and Stubblefield, J. W. (2001). Models of sex ratio evolution. In Hardy (ed.): *Sex Ratios – Concepts and Research Methods*. Cambridge: Cambridge University Press: 2-25.
- Skyrms, B. (1997). Chaos and the explanatory significance of equilibrium: Strange attractors in evolutionary game dynamics. In Bicchieri, Jeffrey and Skyrms (eds.): *The Dynamics of Norms*. Cambridge University Press: 199-222.
- Skyrms, B. (2000). Stability and Explanatory Significance of Some Simple Evolutionary Models. *Philosophy of Science* 67: 94-113.
- Sober, E. (1983). Equilibrium Explanation. *Philosophical Studies* 43: 201-210.
- Weintraub, E. R. (1985). *General Equilibrium Analysis*. Cambridge: Cambridge University Press.
- Woodward, J. (2000). Explanation and Invariance in the Special Sciences. *British Journal for the Philosophy of Science* 51: 197-254.
- Woodward, J. (2003). *Making Things Happen – A Theory of Causal Explanation*. Oxford: Oxford University Press.
- Woodward, J. and Hitchcock, C. (2003). Explanatory Generalizations – Part I: A Counterfactual Account. *Noûs* 37: 1-24.
- Ylikoski, P. (2001). *Understanding Interests and Causal Explanation*. Ph. D. thesis, University of Helsinki. Available from <http://ethesis.helsinki.fi/julkaisut/val/kayta/vk/ylikoski/>