Gravitational Model of Rational Choice:
A Framework for Social Mechanism Theorizing

Kai Quek
Cornell University

January 2005

Paper presented at the Center for the Study of Economy and Society, Cornell University.
Research support from the Cornell Commitment is gratefully acknowledged.

1. Introduction

Theory construction based on causal mechanisms stands as one of the most significant methodological developments in sociology in the last fifteen years. By "causal mechanisms" we refer to the entities and activities organized such that their interactions realize of regular changes from explanans to explanandum; the activities are the producers of change, and the entities, with their specific properties, are the objects that engage in the causally productive activities (Machamer et al. 2000). The theorized entities are typically that of a more fine-grained level of analysis. In this case, mechanisms can be understood as "bits of theory" about entities at a different level (e.g. the individual level) "which serve to make the higher-level theory more supple, more accurate, or more general" (Stinchcombe 1991: 367).

Through the specification of causal interactions and processes, mechanism theorizing gives us a form of social explanation that does more than mere causal statements, or assertions about correlations and necessitations (see Elster 1989; Hedstrom and Swedberg 1998). Significant scientific progress can be – and have been – driven by mechanism theorizing, as is readily discernible in other sciences such as biology and neuroscience.[1] The pleasing properties of mechanism-based

[1] For example, a quick keyword search on "mechanism" and "mechanisms" in the titles and abstracts of the journal *Nature* shows 63 hits for issues published in the last five years (1999-2004). For a theoretical argument, see Craver (2001) which argues that scientific change can be characterized in terms of the piecemeal construction, evaluation and revision of multi-level mechanism schemata.

explanations give us good reasons to believe that advancements in social science can be similarly driven by the discovery and description of social mechanisms.

The social mechanism approach, however, remains incomplete, despite important contributions such as Miller (1987), Elster (1989), Little (1991), Stinchcombe (1991), Hedstrom and Swedberg (1998), and others. Perhaps the most serious absence is that of an overarching framework to guide and discipline social mechanism theorizing. A theory that specifies a social mechanism may approximate a better form of causal explanation; however, the specification of a mechanism in itself is insufficient. What is needed is a systematic consideration of how different mechanisms relate to one another in causing the explanandum. While the importance of this point has been vaguely recognized in the literature – as well as demonstrated by some of the best theoretical works in contemporary sociology – there remains no rigorous proposal as to how this may be done. The result is that even with the specification of social mechanisms, three problems remain salient. The first is the problem of incoherence, whereby although different causal processes are considered, they are not related to one another in a coherent and systematic way. The second – and the more fundamental – is the problem of incompleteness, whereby the theory privileges one or a few social processes by ignoring or missing out other possible causal forces. Because data searching can never be theory-free, the bias transmits to data collection and presentation; the final theory is a partial causal story that fails to address the causal complexities

involved in the production of a social phenomenon.[2] The problem of incompleteness leads to the problem of predictive inaccuracy, since incomplete theories can hardly be expected to function as a solid basis for robust predictions.

The solution is to organize social mechanism theorizing with a methodological model that allows us to systematically analyze the relationships between different causal mechanisms. Many demands will be placed on this framework. Primarily, we require that it assists us in attacking the problems of incoherence, incompleteness and predictive inaccuracy, and in doing so, sharpen the explanatory and predictive potential of social mechanism theorizing. Additionally, we would like the framework to assist us in capturing the complex properties of the social world. In particular, the framework should accommodate the inter-related properties of (i) causal contingency, (ii) probabilistic causation, and (iii) multiple causality. By property (i), a social outcome is obtained as it happens, but the outcome does not *have* to always obtain given the explanans. Whether it will obtain is a probability dependent upon the interaction of prior circumstances, as per property (ii). Property (iii) requires that the framework should allow for the possibility of different causal pathways leading to similar outcomes through different interacting mechanisms. These properties, together with the demands for completeness, consistency and predictive accuracy, are rather reasonable properties to demand of any good theory.

[2] A common example is the class of monocausal theories, which constructs an explanation based on a linear pathway that relates explanans to explanandum in a single causal pattern.

It may appear that our wish-list is somewhat formidable: To surmount the problems of incoherence, incompleteness and predictive inaccuracy which devalue sociological theories, and – on top of that – capture the properties of causal contingency, probabilistic causation, and multiple causality. It turns out, however, that the model developed in this paper may offer a possible solution. An important concept in the framework is the concept of equilibrium. Equilibrium analysis has a distinguished (but sometimes overlooked) history as the foundation stone of economic science. Indeed, it is the analytical pivot in Adam Smith's *The Wealth of Nations*, where he had argued that the position in which market price equals natural price and output equals effective demand is a "centre of repose and continuance" (Smith 1976 [1776]: 75), which will persist unaltered until an exogenous change occurs. One of the main purposes of the present paper is to bring together a rationalist model of social mechanism theorizing with a general form of equilibrium analysis, showing how the analytical power of the latter can be marshaled to strengthen the former. In the process, we will also show how the model may assist us in distinguishing between different types of social mechanisms and their interactive relations with regard to rational action.

The basic idea is to treat social equilibrium and disequilibrium as fully determinable within a model of rational action. Equilibrium, thus defined, will involve rational choice. To specify this conception analytically, we shall develop a gravitational model of rational action that expresses the appeal of rational action in terms of a "gravitational pull" produced in the intersection of the actor's opportunities and interests. Using this model we will consider how opportunities

and interests interact to shape the possibilities of rational action, and how this interaction is in turn determined by different classes of social mechanisms. This will permit a rational-choice formulation of both social mechanisms and social equilibrium that will, together, contribute the properties of completeness, coherence and predictive accuracy, which we desire – as well as the satisfying properties of explanation inherent in the rational-choice paradigm (e.g. Hollis 1977; Coleman 1986; Friedman and Hechter 1988; Boudon 1998).

The paper is organized as follows. In section two, the model is presented. In the process we shall also consider the implications of our model for the concept of rationality, as well as the usefulness of a restricted form of the model. For convenience, three analytic rules-of-thumb for social mechanism theorizing will also be suggested. Since the model introduced is very general, to demonstrate plausible applicability, some quick illustrations using actual sociological cases will be made in section three. Section four concludes.

2. Model

We begin by assuming that causes of social outcomes derive from the actions of social actors (either individual actors or aggregated actors). Actions are either rational or irrational. Only rational action – and the social mechanisms that involve rational action – will be considered here.

Our methodological framework is based on a gravitational model of rational action. The reason for its name will become clear subsequently. Suffice for now to say that the model takes that rationality is to social science what gravity is to physical science. We assume that the human capacity for rational calculation is the universal and systematic force that animates all forms of social interactions, and that the actions of human agents gravitate to the influence of their rational deliberations (just as Newton's apple is influenced by a gravitational field created by the mass of the earth). Of course, rationality is seldom the sole force at work in the production of a social phenomenon; of all the forces, however, it is the most constant and regular, and is probably the closest approximation we have to a universal law in social science, if we should ever desire one.[3]

O-Space, I-Space, R-Space

Define an *opportunity space* (hereinafter "O-space") as the space of all available opportunities for action in a specific issue context. We define an opportunity as a possible combination of circumstantial resources, and each circumstantial resource as being completely specified by its relevant characteristics. The O-space may hence be viewed as the universe of all possible combinations of a finite set of circumstantial resources. For simplicity, we assume that resources are distinguishable into two general classes: endowments and institutional rules.

[3] This of course does *not* mean that it is, without exception, *the* dominant force at work – we know that gravitational force can be offset by other forces. Its candidature as a universal theory lies not so much in its dominance as in its prevalence and regularity. Indeed, the operations of rationality exhibit such regularity and persistence in the social universe that rational choice appears rather deserving of a law-like status as a gravitational law of action – by the more permissive criteria of social science – as does the gravitational law of motion in physics.

We take that each characteristic expresses one distinct dimension of the resource (e.g. monetary cost) which has relevance with regard to the resource's efficacy within the issue context. This will allow us to compare the opportunities along *n* dimensions, where *n* is the number of relevant characteristics. If required, the O-space can be formally conceptualized as an *n*-dimensional Euclidian space, in which each point represents a list of relevant characteristics that defines an opportunity for action.

Define an *interest space* (hereinafter "I-space") as the space of all desired outcomes. Each point in the I-space represents an outcome desired by the actor, which can be interpreted as a particular combination of characteristics. As in our construction of the O-space, we take that each desired outcome is completely specified by its relevant characteristics, and that each characteristic expresses a distinct dimension which is relevant to its desirability. This results in an *m*-dimensional I-space, where *m* is the number of relevant characteristics. Desired outcomes, therefore, are variant along *m* dimensions.[4] For analytical purposes, it is important that the set of desired outcomes is internally consistent. To achieve this aim, various consistency axioms such as that of completeness and transitivity may be imposed; the important point is that no desired outcome contradicts another. At the minimum, with the formal requirements of completeness and

[4] To have an *n*-dimensional O-space intersecting with an *m*-dimensional I-space would present mathematical difficulties that must be addressed when we operationalize the model. The section on "Restricted Model" addresses this issue. For the general model, however, we refrain from imposing assumptions on the number of O-space and I-space dimensions.

transitivity met, we will have at least a weak ordering of preferences.[5] Ideally, therefore, the I-space, bounded in one specific issue context, will express consistent interest orientations with regard to the object of interest.

The *space for rational action* (hereinafter "R-space") is defined by the intersection of O-space and I-space. Each point in the R-space represents a plan to employ an available opportunity in pursuit of a desired outcome. The larger the R-space, the greater the number of rational actions the actor can choose from to attain a certain goal. Since rational action is defined by the intersection of O-space and I-space, an actor is judged to be irrational if he attempts an action that clearly does not reside within the intersection.[6]

Gravitational Model of Rational Action

The *gravitational tendency towards rational action* (denote as G) depends on the size of the R-space (denote as R). The larger the R-space, the greater the number of rational means the actor can choose from to achieve a desired end.

The gravitational tendency towards rational action (G) is proportional to the size of the R-space (R) discounted by a risk-aversion factor (denote as r).

---

[5] Define "$\geq$" as a single relation "preferred or indifferent to." The axiom of completeness requires that for all x and y, either $x \geq y$ or $y \geq x$, while the axiom of transitivity demands that for all x, y, and z, $x \geq y$ and $y \geq z$ imply $x \geq z$. See Arrow (1951: 12-13).

[6] This means we can judge an actor as irrational if he attempts to achieve something he does not want (i.e. out of the I-space), or something that he wants but is unattainable (i.e. out of the O-space).

$$G \sim \frac{R}{f(r)}$$, where r denotes a risk-aversion factor[7]

Since R is resultant of the overlap between O-space and I-space, we denote the size of the overlap between the spaces as $(O \cap I)$, and produce the equivalent expression:

$$G \sim \frac{(O \cap I)}{f(r)}$$

Our formula for the gravitational force of R-space, interestingly enough, turns out to be structurally similar to Newton's law of universal gravitation, which states that every particle in the universe attracts every other particle with a force that is proportional to the product of their masses and inversely proportional to the square of the distance between them. Using the standard notations in physics, Newton's equation is:

$$F = G_z \frac{m_1 m_2}{r_z^2}$$

where F is the gravitational force, m_1 and m_2 are the masses of the two particles, r_z is the distance between them, and G_z is a universal constant. (The subscript z is added to prevent confusion with our own notations.)

[7] The risk-aversion factor can be derived in different ways, depending on the subject of study, and as long as the method of derivation accomplishes logical consistency and the best possible approximation to the specific circumstances studied. It can, for instance, be constructed as an index based on a set of circumstantial indicators.

The structural similarities are interesting to note. In place of a product of masses (m_1m_2), we have an intersection of spaces, (O $\cap$ I). In place of being inversely proportional to the squared distance between the particles, as is Newton's *F*, our *G* is inversely proportional to a function of a risk-aversion factor, $f(r)$. As a consequence, our gravitational model of action in social science becomes conveniently interpretable in a form that is rather analogous to the gravitational model of motion in physics. Newton's *F* directly covaries with the masses m_1 and m_2 of the two particles; our *G* directly covaries with the sizes of O-space and I-space. For Newton, with the force of attraction inversely proportional to the square of the distance, gravitational attraction increases without bound as the distance between the two masses approaches zero. For us, similarly, the gravitational tendency towards rational action increases without bound as $f(r)$ approaches zero – when risk-aversion is zero, gravitation towards rational action becomes absolute, since the actor has no shadow of hesitation.

In our model the O-space and I-space are assumed as independent. That is, changes in the size of the O-space have no impact on the size of the I-space, and vice versa. The resulting interpretation suggests that is not that our interest in a certain goal increases when our opportunities expand, but that the space for rational action expands when opportunities increase, and hence creates a stronger gravitational pull upon us based on our own rational calculations. Note, however, that the independence assumption is imposed merely for the purpose of simplifying analysis; if necessary, it can be removed and the premises correspondingly revised without upsetting our basic results.

Rational Choice

The traditional conception of rationality as self-interest maximization has long been recognized as overly restrictive (e.g. Nagel 1970; Elster 1979, 1983; Hirschman 1982, 1983; Akerlof 1984). Rationality has often been conceived as a deterministic construct and reduced to fixed rule-prescribed behavior. The unhappy consequences include a number of embarrassing paradoxes (e.g. Green and Shapiro 1994) and a serious wedge between rationalist explanations and actually observed behavior.

Rationality is conceptualized in our model as a gravitational pull on action motivated by the reasoning calculus of the actor; a rational action is one which obeys the direction of that gravitational pull. Unlike the traditional conception, the present model is non-deterministic in nature and imposes no severe restrictions on human behavior; there is no requirement to find rationality in unique solution points or tightly-prescribed behavior. In the present model rationality is modeled probabilistically as a gravitational tendency determined by the R-space: the larger the R-space, the stronger its gravitational pull, and the higher the likelihood that rational action will obtain. Since the number of possible combinations of circumstantial resources increases as R-space expands (i.e. more rational options from which to choose), the probability of attaining a desired outcome increases correspondingly as R-space becomes larger.

While a deterministic conception is readily acceptable in a purely normative formulation of rationality (i.e. how we prescribe "rational" actors *should* act), a probabilistic construction is superior in the case of a descriptive model of how rational actors *do* act. If social explanation and prediction are intended – and indeed they should – the latter construction must be judged as preferable. This is because even if we believe that an actor is rational, analytical predictions of his behavior must be expressible in probabilistic terms, as they are *ex ante* in nature. To model rational behavior as a deterministic necessity will be to ask too much of human behavior.

Another useful implication that flows from our model is that self-interest maximization is not required as a necessary condition for rationality. In contrast to the standard model used in economics, the present model is rather forgiving. Maximization, while typically desirable, is unnecessary – non-maximization is not treated as equivalent to irrationality. Furthermore, whereas in the literature the scope of "rationality" has expanded to describe not only actions but also interests and outcomes, the present model returns the focus to the actor's action. Interests or outcomes *in themselve*s are never rational or irrational; only actions are. The view that interests and desires cannot be deemed (ir)rational is not new; it has been made as early as in Hume (1978 [1740]).[8] The point is that the concept of rationality should not be used to pass judgments on what actors want or obtain,

[8] According to David Hume, passions in themselves cannot be called unreasonable or irrational: "a passion must be accompay'd with some false judgment, in order to its being unreasonable; and even then 'tis not the passion, properly speaking, which is unreasonable, but the judgment … a passion can never, in any sense, be call'd unreasonable, but when founded on a false supposition, or when it chuses means insufficient for the design'd end" (Hume 1978[1740]: 416).

but only on what they choose to do. Rationality can only be used to describe interests and outcomes when they are not independently considered. In that case, if it is evidenced that a desired outcome contradicts our axioms of internal or external consistency, we would have sufficient justification to judge it as irrational. However, when the consistency axioms are obeyed, we have no justification to make value judgments on whether the actor should view an object of interest as good or bad.

What the gravitational model expresses is a minimal structure of rationality, whereby a rational actor gravitates towards action within the intersection of O-space and I-space. Internal consistency within the set of desired outcomes is required and – if multiple issue contexts are considered – we may also demand external consistency with the sets of desired outcomes in related issue contexts.[9] The main point is to prevent contradictory or mutually-defeating interests both within as well as across issue contexts. For this purpose, formal axioms of consistency may be imposed as necessary. It is conceivable, of course, that the model may be further specified with more elaborate rules of decision. For instance, as is standard in microeconomics, we may impose a set of choice rules that prescribes maximization along one or a few select dimensions in R-space. Alternatively, we may have non-maximizing choice rules such as those which

[9] There are various plausible ways to formalize external consistency, each with its appeals and pitfalls. The most intuitive way will be to simultaneously submit all interests, irregardless of their issue contexts, to the completeness and transitivity axioms. Another way might be to derive ordinal weights based on a complete rank-ordering of issue contexts, and thereafter to ensure completeness and transitivity among the most salient interests of each issue context on the basis of the ordinal weights. To consider these issues in depth will be the subject of another paper.

model habit-dependent decisions. These elaborations, however, are not necessary conditions for us to define the basic structure of rationality. Indeed, they will only be useful if we have enough empirical information to justify their use in the given issue context.

In sum, what we have done is to express the structure of rationality in its basic form. The appeal of our model is that it is both more and less than standard conceptions of rationality. It gives us more than rationality as internal consistency, which is too weak a demand (e.g. Sen 1982), and it gives us less than rationality as self-interest maximization, which is too strong a demand (e.g. Simon 1955, 1957, 1982). It abandons prescriptive determinacy and gives us a probabilistic tendency. It provides a basic structure necessary for identifying rational behavior, and bypasses the computation of fixed solution points. Rationality is formulated as a causative gravitational force resultant of the actor's reasoning calculus, emergent from the intersection of O-space and I-space, and which satisfies a set of consistency axioms. More than this is not necessary.

Social Mechanisms

Since we begun by assuming that the causes of social outcomes derive from the actions of social actors, it follows that there is a direct correspondence between the magnitude of the gravitational pull of rational action (G), and (rationalistic) social causation. Whether or not a social explanandum obtains will depend causally and directly on the magnitude G. Therefore, to explain how a social

explanandum is obtained through the rational actions of interacting agents, social mechanisms must demonstrate how changes in the gravitational pull of rational action (*G*) obtain. Since *G* is determined by the size of R-space – which is the intersection of O-space and I-space – social mechanism theorizing will involve causal mechanisms that track changes in the sizes of O-space and I-space.

The O-space comprises of two broad categories of circumstantial resources – endowments and institutional rules. There are hence two general classes of O-space mechanisms to consider. First is the class of endowment mechanisms which produce and reproduce the diverse stock of endowments that changes or maintains the O-space. This category consists of various sub-classes of endowment mechanisms, each corresponding to a different type of endowment. For instance, common types of endowment include that of social capital, financial resources, technical knowledge, and so on.

Next is the class of institutional mechanisms which create and enforce the institutional rules that change or maintain the O-space. It might be tempting to assume institutional rules as a standard bundle of resources possessed in common by members of a given society. To do so will oversimplify, however. Although the law may apply to all actors equally, differences would be generated by variations in the combinations of resources across individuals and groups; for example, knowledge of and access to legal means are always unequal. There is also the need to accommodate the possibilities of rule-breaking and institutional loopholes, which will present additional opportunities for action. These possibilities will vary

according to the special endowments available for rule-breaking (e.g. political clout or Mafia connections) and the exploitation of institutional loopholes (e.g. legal knowledge or bribe money).

I-space mechanisms are mechanisms that produce and reproduce the motivating interests that change or maintain the I-space. The I-space consists of desired outcomes, which may refer to physical goods or states of affairs. Interests are derived from (i) intrinsic reasons (e.g. pain avoidance) and (ii) locational reasons, by which we assume that an actor's interest is influenced by his position within a specific pattern of relations with other actors (e.g. through mechanisms such as that of intra-group socialization or conformity). The former implies primitive psychological mechanisms; the latter implies social mechanisms grounded in the specific pattern (or network) of relations, beliefs and expectations between actors. Purposive action is motivated by and embedded in concrete and ongoing systems of social ties, religious and cultural beliefs, customs, and norms (Granovetter 1985; Weber 1992 [1904-5], 1978 [1922]). Locational interests suggests the relativity of interests: An interest and its value are often defined based on where the actor is situated in a specific pattern of relations, as well as the properties of this pattern, such as the character, density, intensity and extensity of patterned relations.

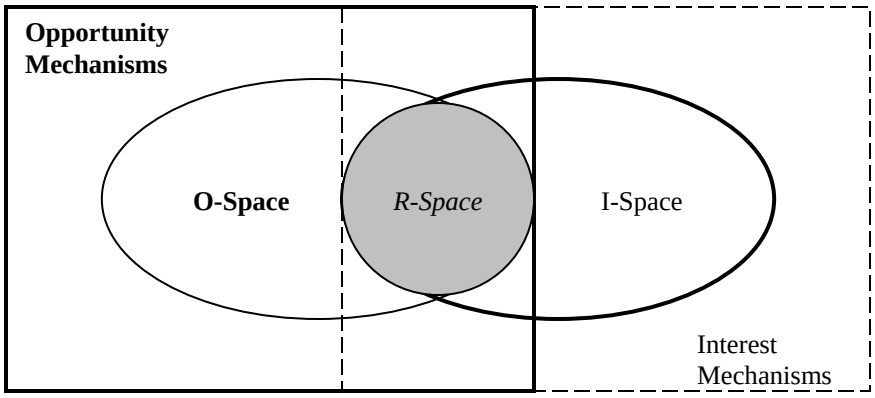

FIGURE 1
Two-dimensional representation of the model

A few additional points are worth mentioning. First of all, when we consider aggregated actors (e.g. groups or organizations), it is useful to take that aggregation implies a certain homogeneity of interests based on either consensus over desired outcomes or a general similarity in interest orientations. This condition is required for the I-space to be well-defined. If interests are significantly divergent and heterogeneous, it might be analytically more meaningful to disaggregate the aggregated unit into separate and distinct entities until sufficient interest homogeneity is achieved. Secondly, although our framework does not explicitly model path dependence, the concept may be easily incorporated: (i) by considering, for I-space, the location of an actor at time *t* as carried over from time *t-1*; and (ii) by considering, for O-space, the inherited endowments and institutional rules carried over from time *t-1* to time *t*. Finally, our model reveals how exactly the problem of incompleteness handicaps a theory. Consider, for instance, the class of monocausal theories that is rather prevalent in

sociology. Such theories are built upon linear pathways which relate explanans to explanandum in a single causal pattern. In effect, when a theory privileges one causal factor, it has to assume that all other factors that may affect the sizes of I-space and O-space are constant or – at least – that the variation in these parameters are so small as to have no significant causal effect. These assumptions are very strong. If not explicitly justified and tested with empirical facts, such assumptions are likely to render a theory seriously incapable of addressing complex social realities.

Social Equilibrium

A general form of equilibrium analysis is exercised in our model to estimate the stability of the arrived outcomes. Clearly, changes in the determinants of I-space and O-space for one distinct configuration of actors in one issue context implicate changes in the determinants of I-space and O-space for a different configuration of actors in a related issue context. Put simply, it means that the causal story may not end within the span of "explanans → (mechanisms) → explanandum" if the derived explanandum is not a stable outcome – that is, if it significantly disturbs the interests of other rational actors. Social changes at one stage may lead to structural changes in the configuration of interests, opportunities and relationships between relevant actors at the next stage, which in turn would lead to rational response and counter-responses, until a stable interest-coincident outcome is obtained. Therefore, to complete our model, it is necessary to consider how the

existence of stable equilibrium outcomes may be analyzed from the linkages between multiple sets of social mechanisms.

There are numerous difficulties in defining a useful conception of equilibrium via the mathematical identification of unique solution points to a particular formal model, as is widely practiced in economics. To escape these complications, we make an opposite projection: we define equilibrium not by what *is*, but by what *isn't*. In so doing, we apply the concept of equilibrium as an organizing category for theory construction, instead of limiting it as a solution concept relevant to a particular model. By our model, the existence of a social equilibrium is defined by the *absence* of R-space – *equilibrium exists when the R-space for countervailing actions is zero.* When there is no intersection between I-space and O-space, the gravitational pull of action (G) is zero, since there is no room for rational action (i.e. no incentives to act). Equilibrium, when achieved, represents an outcome from which there is *no* endogenous tendency to change. Instead of identifying unique *solution points* of equilibrium, our analysis proceeds by gauging the *zone* of equilibrium, which is defined by the gap between the I-space and O-space for countervailing actions. The larger the gap between interests and opportunities, the more stable the equilibrium. Conversely, the extent of social *disequilibrium* is defined by the size of the R-space; the larger the room for rational countervailing action, the stronger the tendency for change. The smaller the zone of disequilibrium, the more stable the outcome, and the smaller the likelihood of a countervailing change.

The appeals of this formulation are at least two. Firstly, theoretical notions of equilibrium sometimes involve the strong assumption that there is concurrently a state of absolute rest for all other factors, as in Alfred Marshall's (1920[1890]) famous ceteris paribus condition; or at least that all possibly related variables are constant. The present model proposes an alternative formulation which reduces our need for heroic assumptions. In our rationalist framework, all that is relevant is that O-space and I-space do not intersect; when they do, the resulting R-space translates into a positive likelihood of change.

Additionally, we should note that the economic concept of equilibrium faces at least two problems when applied to social analysis. First, the data requirements needed to compute solution points for social processes may be infeasible. Second, the models themselves may be fundamentally inaccurate, being typically characterized by tight and unrealistic assumptions. Computing equilibrium solutions – although mathematically pleasing – is substantively empty when the assumptions of its model are deeply divergent from reality. In contrast, the mere requirement of testing for '0's to arrive at equilibrium judgments is far less demanding. Empirical evidence is also more readily attainable than is the case for calculating and empirically proving the exact solution points in a model. So while our formulation is broad, it also requires less. The empirical demands necessary for us to make equilibrium judgments are reduced to a more feasible level.

Restricted Model

Presented in the preceding sections is our general model. The general model provides a heuristic framework for social mechanism theorizing, which is the aim of the present paper. It is apparent, however, that the model may also be useful for middle-range modeling of specific social phenomena. In this section we shall digress to comment very briefly on how this may be done.

For specific empirical application, a restricted form of the general model should be constructed. The main purpose of the restricted model is to impose a set of assumptions to reduce the number of dimensions in O-space and I-space, so as to simplify analysis and to be more specific to the character of the phenomenon analyzed. For instance, assumptions may be devised to approximate the cognitive limitations of actors in different choice situations, or to focus analytical attention on a specific list of critical endowments. Or it may be the case that a certain choice rule (for instance, profit maximization, local status maximization [Frank 1985], or satisficing behavior [Simon 1982]) is assumed because it is highly probable in a given context. [10]

For illustration, let us formulate an assumption which is quite universally useful, and which arrives from the observation that the O-space and the I-space may share many common dimensions. This observation should not be surprising – recall that I-space is the space of all desired outcomes, and that an outcome is a product of an opportunity. Using this insight, let us treat a point in I-space as an *imagined*

[10] Note that the question of which exact point to choose in the R-space can be answered when we impose a determinate set of choice rules. The issue has been left open in the general model, since there is no need to fix a determinate choice rule for the methodological framework to work.

opportunity completely specified by its desired characteristics. Thus defined, a desired outcome becomes interpretable as a desired opportunity to achieve an outcome. With some reductionistic assumptions, we can ensure that I-space differs systematically from O-space in only two dimensions: (i) the desire-intensity dimension and (ii) the availability dimension. For (i), every point in I-space will be subject to characterization in terms of its desire intensity; on the other hand, every point in O-space will have a neutral value of '0'. For (ii), because desired outcomes are "imagined", every point in I-space will have a value of '0' in this dimension. By this formulation, we force the *n*-dimensional I-space and *m*-dimensional O-space into the exact same number of dimensions. This allows us to integrate them formally into an *n*-dimensional Euclidian space, in which every point will capture a list of fully comparable characteristics.

A key advantage in working with a restricted model is that it enables easier operationalization. The simplest method is to construct a restricted model that emphasizes only one critical characteristic dimension in each of the O-space and I-space for which there are accurate indicators. Consider, as an illustration, the dimension of monetary cost. Suppose that the issue context allows the relevant resources in O-space to be approximated in monetary cost terms. We may then fix a cost axis along which we may enumerate and compare all the possible opportunities (or the most salient ones).[11]

[11] As a practical example, consider how public policy may manipulate the O-space to encourage or deter various forms of rational action through manipulating the monetary costs of action (e.g. through taxes and fines). A diagrammatic illustration of an effective enlargement of the O-space through changing institutional rules (e.g. tax incentives) can be found by contrasting Figures 2A and 2B.

Consider for I-space the critical characteristic dimension of desire intensity. Let us assume that the greater the desire intensity, the greater the costs which the actor is willing to incur in order to achieve the object of interest. This assumption allows us to collapse the desire-intensity dimension into the cost dimension. Suppose the issue is to obtain a certain state of affairs *S*. Our restricted model may treat *S* as a good with only one fixed combination of characteristics (or that the actor is indifferent to the varying combinations of characteristics insofar some minimal threshold is met for him to define the state of affairs as *S*). What matters, then, is *S* plus the cost dimension. If monetary cost provides reasonable approximations, then the I-space is reducible to a one-dimensional continuum of $S + \$p$ for which $\$0 \leq \$p \leq \$q$, where $\$q$ is the maximum monetary cost the actor is willing to pay. The continuum is larger if the desire intensity is high, and smaller if otherwise. In this very simple restricted model, the intersection of O-space and I-space is collapsed into a single dimension – a monetary cost axis. Of course, this formulation is rather crude; for more complicated cases, more than one critical characteristic dimension will have to be considered – but there will be no serious difficulty if there are suitable indicators for each of the dimensions considered. The idea is the same.

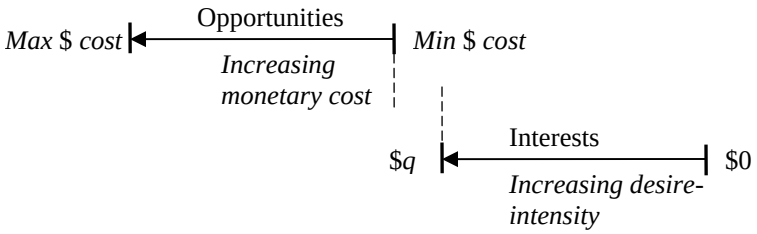

FIGURE 2A
One-dimensional restricted model
in which R-space is non-existent

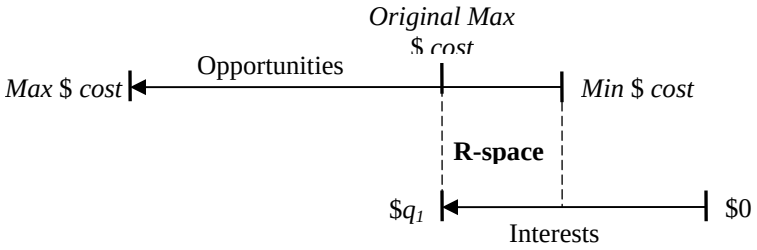

FIGURE 2B
R-space generated by the enlargement of O-space
Enlargement occurs through the creation of lower-cost opportunities

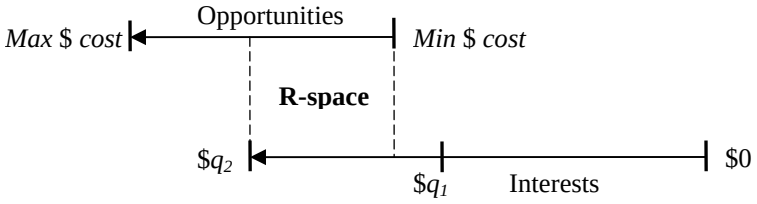

FIGURE 2C
R-space generated by the enlargement of I-space
Desire intensity increases; the maximum cost the actor is willing to incur increases from $q1 to $q2

Analytic Rules

From our general model, we can summarize three analytic rules (AR) as a convenient guide for social mechanism theorizing:

AR(1) Check for the possible existence of the R-space (and hence the gravitational tendency towards action) by checking for changes in the size of I-space and O-space. This requires that we test for possible mechanisms which may influence the determinants of I-space (specific patterns of relations) and O-space (endowments and institutional rules).

AR(2) Check whether the theorized social mechanisms are by themselves necessary and sufficient to produce the explanandum. This involves checking that other relevant social mechanisms that may affect the size of I-space or O-space (e.g. those that are known to produce significant effects in a different time or place) are either constant, or that they are so small as to exercise no significant causal effect. This check should be done by testing the empirical implications that flow from the social mechanisms.

AR(3) Check for the existence of social equilibrium. If a theory results in an explanandum that is clearly unstable, the implications of the disequilibrium should be further explored.

3. Illustration

To illustrate our methodological framework, we shall present three quick explanatory sketches based on actual sociological studies. Alba and Nee's (2003) work on assimilation and Perry's (1980) work on peasant rebellion will serve as positive examples, while a quick glance at the history of alcohol prohibition in the US will reveal a negative example. Because all of these issues are complex, we shall strip the issues to their bones and consider – very briskly – only the basic structure of their theories.

Alba and Nee (2003)

Alba and Nee (2003) is an example of a robust theory generated by social mechanism theorizing, and which is consistent with the framework proposed here. To illustrate, we apply the analytic rules AR(1) – (3) as criteria to assess – very briefly and summarily – the theory in Alba and Nee (2003). For assimilation to be a rational outcome, the R-space for assimilation must be clearly existent. The gravitational pull of assimilation for rational actors is determined by the size of the intersection between I-space and O-space. For assimilation to occur there must have been significant changes in the size of I-space and O-space over time.

By AR(1), we check for the different classes of social mechanisms which may influence the determinants of I-space and O-space: Specific patterns of relations

between actors, as well as the stock of endowments and institutional rules. In Alba and Nee (2003), the O-space is expanded by changes in institutional rules and endowments. The institutional structure – sustained both by formal rules and informal constraints – begun a fundamental shift with the Civil Rights movement, which deeply altered the cost-feasibility and practicability of discrimination. However, this expansion in O-space, although necessary, is insufficient to explain assimilation – the O-space is expanded by the same extent for every ethnic group, but there are differences in the extent of assimilation across the groups. The differences are attributed to variations in endowments across groups, which further expand the O-space for the different groups by varying degrees. Alba and Nee (2003) considers the class of mechanisms affecting endowment in a "mix-of-capital" model, whereby different mixes of capital possessed by immigrant families contribute to different outcomes of social mobility and assimilation. Some immigrant groups assimilate at a higher rate because their O-space is larger due to greater endowments of economic and social capital, as is empirically demonstrated.

The theory also captures the possible variations in I-space. The individual's interest to succeed in the American society clearly affects the size of the I-space as theorized in our model. The network of relations in which the individual finds himself affects his locational interest towards economic success. If, for instance, one is socially tied to a number of significant others who exert high expectations of success, it is conceivable that his interest to succeed will be more intense. This affects the size of the I-space with regard to assimilation, since individuals'

purposive action to improve their probabilities of success in the American mainstream leads to unintended consequences that contribute to assimilation. In sum, some groups assimilate more than others due to their larger I-space and O-space, which increase the gravitational pull of their rational action towards assimilation.

AR(2) requires that we check if the repertoire of mechanisms in Alba and Nee (2003) are by themselves necessary and sufficient to produce assimilation. One general determinant which affects the size of O-space is not explicitly considered: Technologies. This means that the theory has implicitly assumed that the change in O-space generated by technological changes is relatively minor. It seems that the assumption is relatively unproblematic. In the period considered, technological change does not seem to exercise a critical causal effect on the assimilation process for the different groups. It would have been interesting, of course, if the effects of the new information, communications, and transportation technologies over time are also considered and tested, since they might have altered the density and intensity of cross-group interactions. However, from the empirical record, it appears that Alba and Nee's (2003) repertoire of mechanisms are by themselves sufficient to parsimoniously account for assimilation, as well as its varying rates across different groups.

Finally, AR(3) requires that we check for the existence of social equilibrium. In Alba and Nee (2003), assimilation is a stable equilibrium outcome insofar it does not deeply revise the I-space and O-space of important actors (ethnic groups) in

other important issue contexts (such as major economic or political issues). As it stands today, in general, assimilation does not significantly alter the configuration of interests and opportunities for other actors in other key issue contexts. There is, therefore, no sizable R-space (no strong gravitational pull) for rational actors to purposively obstruct assimilation in the US. The explanandum is therefore a rather stable equilibrium, as it stands now.

Alba and Nee's (2003) theory manages to unify a parsimonious set of I-space and O-space mechanisms and to show how the operations of these mechanisms vary across different groups to shape different trajectories of assimilation. Assimilation is theorized as a rational outcome contingent upon the joint effect of individual choice and collective action within an institutional structure. Empirical facts are marshaled to support the theorized repertoire of mechanisms. By satisfying AR(1) to AR(3), the theory is shown by our framework to have successfully surmounted the problem of incompleteness which obstructs successful social mechanism theorizing; it has also achieved a theoretical structure that accommodates the desirable properties of causal contingency, probabilistic causation and multiple causality.

A Theory of Rebellion and a Policy of Prohibition

Alba and Nee (2003) was applied as a compact sketch of our framework, and we shall not involve ourselves to illustrate the same point three times. The next two examples will focus on flashing out some key elements in our framework.

Perry's (1980) theory of rebellion applied to the Nien uprising in Qing China provides an example to briefly illustrate the workings of AR(3), i.e., social equilibrium analysis. Let us strip the explanation to its basic causal sequence. The first stage is the confluence of adversarial factors (i.e. disturbance to the means of livelihood created by the conjunction of natural disasters, inefficient governance, increased taxes, Taiping rebels, and the personal ambitions of government officials and bandit leaders). Opportunities for survival through traditional, legal means became diminished; on the other hand, opportunities for survival through illegal activities became larger as bandit gangs became stronger and more prevalent. Membership in bandit gangs became attractive to desperate peasants as a rational recourse. Because this outcome threatened the interests of vulnerable villages and the local authorities, it was unstable, and had generated a rational response in the form of greater militia efforts. Conflict escalated between militias and bandit gangs, and bandit gangs – strengthened by more desperate peasants and many surrendering villages – became so large that it began to threaten the stability of the Manchu regime. Previously inactive, the central government intervened, causing the institutional environment to change drastically. This institutional change severely contracted the O-space for the bandits; to continue as before was no longer a rational action. A rational response had to be devised. The bandit gangs – to survive – pooled together their resources (which expanded their O-space for effective resistance) and organized themselves into a rebellion.

What we see in this thumbnail sketch is that changes in the determinants of I-space and O-space (e.g. endowments, institutional rules and pattern of relations) for one configuration of actors in one issue context implicate changes in the determinants of I-space and O-space for another configuration of actors in another issue context. Social changes at one stage lead to structural changes in the configuration of interests and relations between different groups at the next stage, leading to a series of responses and counter-responses until equilibrium is restored (i.e. when the Nien rebellion was finally crushed in 1868). This short illustration shows that robust theorizing must consider the existence of stable equilibrium outcomes by examining the linkages between multiple sets of social mechanisms that affect the I-space and O-space of important actors in related issue contexts. A theory is regarded as complete and circumspect only when it explains the explanandum as a stable outcome from which there is no strong endogenous tendency to change.

The utility of our methodological model, however, may be extended beyond theory – it may, for instance, have significant value for policy analysis. Indeed, a social policy that significantly diverges from AR(1) – AR(3) must be one that is so badly "theorized" that it is almost certain to fail. A negative example makes the point. One of the greatest social experiments in US history is the policy of national prohibition. There are various interpretations on why prohibition had failed. A quick explanatory sketch, however, can be produced using AR(1) – AR(3), with which it would have been clear that how prohibition was implemented would almost certainly lead to serious problems. For instance,

alcohol consumers tend to have inelastic demand, which translates into a rather rigid I-space.[12] To eliminate the gravitational pull of alcohol consumption as a form of rational action, the O-space has to be severely contracted. A change in institutional rules is insufficient; there must be enough enforcement. However, that was impossible since the amount of resources needed to enable credible enforcement across the nation was simply too much. On the other hand, the prohibition, by driving out all legal brewers and bars (and vacuuming out the competition), had expanded the O-space for organized crime syndicates. History showed that the gravitational pull of lucrative rational action was irresistible, as the syndicates performed the necessary smuggling and running of speak-easies to take over the multi-billion dollar business.[13] As a consequence, the syndicates prospered with greatly expanded economic, organizational and labor resources. When prohibition ended, these syndicates could not be rationally expected to disband their businesses happily. Indeed, the O-space for organized crime syndicates had expanded so much with their enhanced amounts of endowments that institutional control was unlikely to retain the same power over them. With their larger O-space – and insufficient government enforcement to rein in the opportunities – one could only expect rational gravitation toward the lucrative big businesses of illegal gambling, drug-smuggling, and prostitution. Flunking AR(1)

[12] Lusk (1932: 46) went as far as to argue that "National prohibition has made no noticeable difference in the amount of alcoholic liquor consumed … indeed, the absolute amount of alcohol that is being drunk is probably greater …." Lusk's (1932) postulated that liquor consumption had not changed materially; what had changed was the manner in which it was used, those who used it, and the kind of liquor being drunk (e.g. home-brew and adulterated Jamaica ginger).

[13] A police commissioner of New York City claimed that in 1929 there were 32,000 speak-easies in the city (Sanford 1932: 43).

to AR(3), it appears that the policy of national prohibition had achieved not a more teetotaler nation, but stronger mobs and higher crime rates.

4. Conclusion

It appears that our rather parsimonious framework is potentially capable of application to a variety of practical and theoretical areas. Based on a simple gravitational model of rational action, the framework appears useful in disciplining social mechanism theorizing into greater completeness, coherence and predictive accuracy. The framework also reflects the properties of causal contingency, probabilistic causation and multiple causality – all of which are desirable for the accurate modeling of social reality. In addition, we have also briefly extended our theoretical considerations to two further possibilities: (i) the use of the model to express the minimal structure of rationality; and (ii) the use of a restricted form of the model as a middle-range model of specific social phenomena. These theoretical possibilities may be of independent interest, even aside from the methodological argument presented in this paper.

Social mechanism theorizing holds great promise for the advancement of social science; however, without an overarching methodological framework, it is difficult to ensure a complete and systematic consideration of how different mechanisms relate to one another. In this paper we considered how a gravitational model of rational choice, when employed as a framework for social mechanism theorizing, might provide a coherent organizing structure for the analysis of

complex social phenomena. It is hoped that social mechanism theorizing within such a framework might provide significant analytical leverage for both theory and policy.

References

Akerlof, G. (1984). *An Economic Theorist's Book of Tales*. Cambridge: Cambridge University Press.

Alba, R., & Nee, V. (2003). *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. Cambridge, MA: Harvard University Press.

Arrow, K. J. (1951). *Social Choice and Individual Values*. New York: Wiley.

Boudon, R. (1998). Social Mechanisms without Black Boxes. In P. Hedström & R. Swedberg (Eds.), *Social Mechanisms: An Analytical Approach to Social Theory* (pp. 172-203). Cambridge, UK; New York: Cambridge University Press.

Coleman, J. S. (1990). *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.

Craver, C. (2001). Role Functions, Mechanisms, and Hierarchy. *Philosophy of Science, 68*(1), 53-74.

Elster, J. (1979). *Ulysses and the Sirens: Studies in Rationality and Irrationality*. Cambridge: Cambridge University Press.

Elster, J. (1983). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press.

Elster, J. (1989). *Nuts and Bolts for the Social Sciences*. Cambridge, UK: Cambridge University Press.

Frank, R. (1985). *Choosing the Right Pond: Human Behavior and the Quest for Status*. New York: Oxford University Press.

Friedman, D, & Hechter, M. (1988). The Contribution of Rational Choice Theory to Macrosociological Research. *Sociological Theory, 6*, 201-218.

Granovetter, M. (1985). Economic Action and Social Structure: The Problem of Embeddedness. *The American Journal of Sociology, 91*(3), 481-510.

Green, D. P., & Shapiro, I. (1994). *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*. New Haven: Yale University Press.

Hedström, P., & Swedberg, R. (Eds.). (1998). *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge, UK; New York: Cambridge University Press.

Hirschman, A. O. (1982). *Shifting Involvements*. Princeton, NJ: Princeton University Press.

Hirschman, A. O. (1983). Against Parsimony: Three Easy Ways of Complicating Some Categories of Economic Discourse. *American Economic Review, Papers and Proceedings, 74*(2), 89-96.

Hollis, M. (1977). *Models of Man: Philosophical Thoughts on Social Action*. Cambridge: Cambridge University Press.

Hume, D. (1978 [1740]). *A Treatise of Human Nature*. Oxford: Clarendon Press.

Little, D. (1991). *Varieties of Social Explanation: An Introduction to the Philosophy of Social Science*. Boulder, CO: Westview Press.

Lusk, R. S. (1932). The Drinking Habit. *Annals of the American Academy of Political and Social Science, 163*, 46-52.

Machamer, P. K., Darden, L., & Craver, C. F. (2000). Thinking About Mechanisms. *Philosophy of Science, 67*, 1-25.

Marshall, A. (1920 [1890]). *Principles of Economics*. London: Macmillan.

Miller, R. (1987). *Fact and Method: Explanation, Conformation and Reality in the Natural and the Social Sciences*. Princeton, NJ: Princeton University Press.

Nagel, T. (1970). *The Possibility of Altruism*. Oxford: Clarendon Press.

Perry, E. J. (1980). *Rebels and Revolutionaries in North China, 1845-1945*. Stanford, CA: Stanford University Press.

Sanford, E. P. (1932). The Illegal Liquor Traffic. *Annals of the American Academy of Political and Social Science, 163*, 39-45.

Sen, A. (1982). *Choice, Welfare and Measurement*. Cambridge, MA: MIT Press.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics, 69*(1), 99-118.

Simon, H. A. (1957). *Models of Man*. New York: Wiley.

Simon, H. A. (1982). *Models of Bounded Rationality*. Cambridge, MA: MIT Press.

Smith, A. (1976 [1776]). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford: Oxford University Press.

Stinchcombe, A. L. (1991). The Conditions of Fruitfulness of Theorizing About Mechanisms in Social Sciences. *Philosophy of the Social Sciences, 21*, 367-388.

Weber, M. (1978 [1922]). *Economy and Society: An Outline of Interpretive Sociology* (E. Fischoff, Trans.). Berkeley: University of California Press.

Weber, M. (1992 [1904-5]). *The Protestant Ethic and the Spirit of Capitalism* (T. Parsons, Trans.). London: Routledge.