Methodological Individualism, Explanation, and Invariance

Received May 18, 2005

Daniel Steel

Department of Philosophy
503 S. Kedzie Hall
Michigan State University
East Lansing, MI 48824-1032

Email: steel@msu.edu
Phone: (517) 355-4490
FAX: (517) 432-1320

Acknowledgements

I would like to thank an audience at the 2005 Philosophy of Social Science Roundtable for stimulating a helpful discussion of an earlier draft of this essay. I am also grateful for helpful comments and suggestions received from two anonymous referees.
Methodological Individualism, Explanation, and Invariance

Abstract

This essay examines methodological individualism in terms of the theory that invariance under intervention is the signal feature of generalizations that serve as a basis for causal explanation. This theory supports the holist contention that macro-level generalizations can explain, but it also suggests a defense of methodological individualism on the grounds that greater range of invariance under intervention entails deeper explanation. Although this individualist position is not threatened by multiple-realizability, an argument for it based on rational choice theory is called into question by experimental results concerning preference reversals.

Keywords: methodological individualism, mechanisms, explanation, invariance, preference reversal.

Introduction

Methodological individualism is commonly understood as the proposition that social phenomena are best explained in terms of the motivations and interactions of individual persons. Assessing such a claim would presumably require a general understanding of what genuine scientific explanations are and what makes some better than others. Yet surprisingly, the vast literature on methodological individualism makes only sparse reference to philosophical accounts of explanation. In this essay, I take some steps towards remedying that situation. In particular, I wish to argue for the following four theses.
Methodological Individualism, Explanation, and Invariance

(1) Since methodological individualism is generally understood as a thesis about explanation, it can only be properly evaluated within the context of a well-motivated theory of what explanation consists of.

I examine methodological individualism from the perspective of Jim Woodward’s (2003) theory of causal explanation, which is one of the leading proposals on this topic in the current philosophy of science literature. According to Woodward’s theory, the distinction between mere correlations and generalizations that can support causal explanation is that the latter are invariant under some range of interventions: the greater the range of invariance, the deeper the causal explanation. This leads to the next thesis.

(2) Woodward’s theory of causal explanation supports the holist assertion that group level generalizations can explain, but it also suggests an argument that individual level accounts are required for deep explanation.

That is, it seems clear that some collective or group level generalizations in social science are invariant under some interventions. Nevertheless, the individualist could argue that the generalizations involved in individual level explanations are invariant under a significantly broader range of interventions than their group level counterparts. If this were true, Woodward’s theory would entail that, although group level generalizations can explain, they cannot match the depth of individual level explanations. In order to examine the cogency of this reasoning, I articulate the concept of a structure altering intervention. This concept is the basis for the argument for the last two theses.

(3) The claim that individual level accounts provide deeper explanations in Woodward’s sense is not undermined by multiple-realizability.
Methodological Individualism, Explanation, and Invariance

Since the multiple-realizability argument is one of the most common arguments given against methodological individualism, I regard (3) as an interesting point. However, methodological individualism as construed here may still be false even if it is not undermined by multiple-realizability. Woodward (2000, 220-221) points out that the “Lucas critique”\(^1\) can be interpreted as an argument for the claim that rational choice models are more robustly invariant under intervention than macro-level generalizations. If sound, this argument would constitute a basis for the version of methodological individualism discussed in this essay. But there are reasons for skepticism.

(4) Experimental results concerning preference reversals call into question the claim that rational choice models are invariant under a broader range of interventions than macro-level generalizations.

Preference reversal experiments raise deep and currently unresolved questions about the extent to which preferences are constructed in the process of decision making rather than stable causes of decisions. Moreover, any argument that individual level mechanisms are invariant under a broader range of interventions would depend on premises about the nature of preference. Thus, the correctness of the version of methodological individualism discussed here depends on empirical questions that have yet to be answered.

Methodological Individualism and Explanation

Since the expression “methodological individualism” has been used to mean a variety of different things, it is important to specify the sense intended here. Lars Udehn (2001) provides a very thorough analysis of methodological individualism from its

\(^{1}\) For discussions of the Lucas critique, see (Lucas 1976; Hoover 1995).
Methodological Individualism, Explanation, and Invariance

historical roots to its current manifestations. According to Udehn, the thesis most commonly associated with methodological individualism is, “Social phenomena must be explained in terms of individuals, their physical and psychic states, actions, interactions, social situation and physical environment” (p. 354).² This statement covers several individualist positions that disagree about the interpretation of one or more of its key terms. For example, Udehn distinguishes strong from weak methodological individualism (pp. 346-349). Strong methodological individualism would insist that the “social situations” mentioned in the statement above are themselves explicable by individualist means. In contrast, weak methodological individualists allow that there are autonomous institutions and social structures that shape individual behavior. The distinction between strong and weak methodological individualism, however, is not the concern here. Rather, my focus is upon the “must” and especially the “explained.”

The claim that social explanations must be stated via individualist concepts could naturally be read as the claim that only individual level accounts provide any explanation in social science at all. However, such an extreme claim is not easy to defend, and even some of the most ardent advocates of methodological individualism grant that explanations relying upon group or collective level concepts can explain to some extent. Such individualists nevertheless insist that such explanations are superficial, and that individual level detail is required for deep explanation. An example of this is J. W. N. Watkins’ claim that, although one might give “unfinished or halfway” explanations by way of collective social features, no explanation of a social phenomenon has reached

² For example, according to Elster, methodological individualism asserts that, “To explain social institutions and social change is to show how they arise as the result of the action and interaction of individuals” (1989, 13). In Popper’s formulation, methodological individualism “insists that the ‘behavior’ and the ‘actions’ of collectives, such as states or social groups, must be reduced to the behavior and to the actions of individuals” (1966, 91).
“rock-bottom” until given in individualist terms (1957, 105). A similar sentiment is expressed by Jon Elster (1985, 6), and I think that there is a good case for attributing a view of this kind to Mill.

But the claim that individual level explanations are deeper than collective level ones can itself be interpreted in more than one way. On one reading, the “deepness” could be understood to indicate that individual level explanations are always superior to explanations couched at the level of collective social properites. Phillip Pettit refers to this attitude as the “fine-grain preference,” a position for which he argues there is not much to be said (1993, pp. 253-257). Pettit points out that we often desire comparative information from an explanation and that demonstrating common trends across a variety of situations requires abstracting from the nitty-gritty details of individual motivations and interactions. In a similar vein, other authors have maintained that the multiple-realizability of social types entails that explanations provided in terms of collective or group level concepts will often be more unified than individualist explanations (cf. Jones 1996; Kincaid 1996, chapter 5). Thus, social explanations phrased in terms of collective concepts are often preferable when broad scale comparisons are desired.

I think that the argument alluded to in the foregoing paragraph provides a good reason for rejecting the claim that individual level explanations are always superior in every respect to ones phrased in terms of collective concepts. However, the alleged deepness of individual level explanations need not be interpreted as making a claim of this sort. Instead, the claim could be that individual level explanations possess some special explanatory virtue, or generally possess it to a higher degree than do collective level explanations. That is, even if individual level explanations are not always superior
Methodological Individualism, Explanation, and Invariance

in every respect, there still might be some sense in which they are deeper. But what might this special virtue be? One suggestion is that the supposed deepness consists of providing more detailed information about underlying mechanisms. But the explanatory interest of information concerning micro-mechanisms does not suffice to demonstrate that they provide deeper explanations. For instance, although a critic of methodological individualism, Harold Kincaid grants that “the important questions and the relevant kinds of answers for complete social explanation must make some reference to mechanisms involving individuals” (1986, 511). In other words, some detail about individual motives and interactions are typically part of what people would like to learn when they ask for an explanation of some social phenomena. However, given Kincaid’s proposal, micro and macro-level explanations would be equal partners, since comparative questions involving collective social categories are also part of what people would usually like to learn from a social explanation. So, the question remains: what is it about micro-level accounts that endows them with greater explanatory depth?

The question of whether individual level explanations are deeper than group level explanations in some important sense can only be adequately addressed in the context of a well motivated theory of explanation. In general, since methodological individualism is a claim about explanation, both what it means and whether one should accept it depends in part upon criteria for what a satisfactory explanation is. Yet the literature on methodological individualism says surprisingly little about theories of explanation. For example, Udehn’s nearly exhaustive analysis of the methodological individualism literature refers only to Carl Hempel’s deductive-nomological model of explanation (2001, 355-356). Udehn notes that it is difficult to apply the deductive-nomological
Methodological Individualism, Explanation, and Invariance

model to social science, since there are few, if any, examples of universal, exceptionless social science laws. Although Hempel (1965) did propose a statistical version of the covering law model, there is a more fundamental reason against adopting his model of explanation, namely, there are excellent grounds for judging it to be inadequate. The story of the rise and fall of the covering law model is well told by Wesley Salmon (1989), and I shall not reiterate it here. The relevant point for the present purposes is that the covering law model is not the best philosophy has to offer when it comes to a theory of explanation.

After the demise of the covering law model, a number of philosophers proposed causal theories of explanation, which provided straightforward solutions to the counter-examples raised against Hempel’s proposal. One of the most well known of these is Salmon’s (1984) theory, according to which scientific explanation of an event consists in tracing the causal mechanisms that led up to it (pp. 274-275). Thus, Salmon’s theory might seem to support the claim that individual level mechanisms are required for explanation in social science (cf. Elster 1989, pp. 4-5). However, upon closer inspection, it can be seen that this is not the case, since Salmon’s theory fails to indicate the level of detail at which a mechanism should be described. For instance, critics of methodological individualism have pointed out that, if micro-mechanisms are the sine qua non of explanation, then there is no apparent reason why social science should stop at individuals rather than, say, neuroscience (Kincaid 1996, 181-182; Jones 1996, 125). Conversely, an explanation framed in terms of competition among organizations, though not describing interactions among individual persons, surely provides information about causal mechanisms.
A second theory of causal explanation is due to David Lewis (1986). Lewis proposes that “to explain an event is to provide some information about its causal history” (p. 217). Lewis proposes a counterfactual theory of causation, which asserts that $c$ caused $e$ exactly if there is a sequence of events from $c$ to $e$ such that each event depends counterfactually on the one before it (1986, 167). An event $b$ depends counterfactually on $a$ just in case had $a$ not occurred, $b$ would not have occurred either. It is clear that Lewis’ theory of causal explanation provides no comfort to the methodological individualist, since there can be chains of counterfactual dependence among events characterized at the collective level. Approaches to explanation that do not take causation as their touchstone also fail to support methodological individualism. For example, theories that emphasize unification (cf. Kitcher 1989) would support the holist’s claim that abstracting from the details of causal mechanisms can increase explanatory power. Similarly, pragmatic approaches to explanation (cf. van Fraassen 1980, chapter 5), which emphasize the role of the conversational context in which the request for explanation is posed, provide no support for methodological individualism.

So, is there any well motivated theory of explanation capable of suggesting an explanatory virtue or power possessed by only individual level accounts of social phenomena, or at least possessed by them to a greater degree? In the next section, I describe a theory of explanation that does suggest an argument for the claim that individual level explanations are deeper than those given at the collective level.
Methodological Individualism, Explanation, and Invariance

Causal Explanation and Invariance

In a series of articles and a recent book, Jim Woodward has argued for a theory of causal explanation whose centerpiece is a concept he dubs *invariance* (cf. 1999, 2000, 2003). A classic problem in discussions of causation is identifying what feature distinguishes causal generalizations from mere correlations. For example, smoking is statistically associated with lung cancer and barometer readings are statistically associated with the occurrence of storms. Yet only the first of these two relationships is thought to be causal, a fact reflected in the intuition that smoking can explain lung cancer but barometer readings do not explain storms. Woodward proposes that the distinguishing feature is that causal generalizations are invariant under some range of interventions while non-causal, mere correlations are not. In order to qualify as explanatory, a causal generalization need not be invariant under *all* conceivable interventions, but it must be invariant under *some* (cf. Woodward 2003, 251).

The concept of an intervention can be nicely represented by directed graphs, in which arrows represent the relationship of direct cause. In this format, the causal relationships in the smoking-lung-cancer and barometer-storm examples can be depicted as follows:

![Diagram](image)

Figure 1: Prior to the intervention.

In the graph on the left, \(S\) and \(L\) are variables indicating smoking and lung cancer, respectively. In the graph on the right, \(B\), \(A\), and \(T\) are variables representing barometric...
Methodological Individualism, Explanation, and Invariance

reading, atmospheric pressure, and storms. An ideal intervention would be represented graphically in these cases as follows:

![Diagram](image)

\( I \) is an exogenous and is a direct cause only of the variable it targets. In the barometer case, the intervention eliminates the influence of \( A \) upon \( B \), which is represented by the deletion of the arrow. However, it makes no other changes to the causal relationships among \( \{B, A, T\} \). For instance, the influence of atmospheric pressure upon storms persists as before. In the smoking example, there are no arrows pointing into \( S \), so no arrows are deleted. Notice that after the intervention, the link from \( S \) to \( L \) persists, while the connection between \( B \) and \( T \) is severed. That is why the correlation is invariant in the former case but not the latter.

Of course, not all real world interventions are as neat and perfect as this. Since a particular sort of non-ideal intervention will play an important role in the ensuing discussion, a formal definition of an ideal intervention will be useful. Let \( V \) be the set of variables of concern. The notion of an ideal intervention can then defined as follows.

**Definition of Ideal Intervention:** \( I \) is an (ideal) intervention on \( X \in V \) if and only if it is a direct cause of \( X \) that satisfies these three conditions:
Methodological Individualism, Explanation, and Invariance

(i) $I$ eliminates other influences upon $X$ but otherwise does not alter the causal relations among $V$,

(ii) $I$ is a direct cause of no variable in $V$ other than $X$, and

(iii) $I$ is exogenous with respect to $V$.

The intervention is exogenous with respect to $V$ just in case it is neither an effect of any variable in $V$ nor shares a common cause with any variable in $V$. Intuitively, exogenous causes come from “outside” the system. Note that (i) through (iii) are satisfied in the two interventions depicted in figure 2.

The requirement in (i) that an ideal intervention does not change the causal relationships among $V$ besides eliminating influences of $X$ is sometimes dubbed modularity (Hausman & Woodward 1999, 546-7). Modularity is not explicitly incorporated in Woodward’s definition of intervention (2003, 98), but it is implicit in applications of his definition (cf. Woodward 2003, 103), and is explicit in other discussions of interventions (cf. Pearl 2000, 22-3). The interest in modularity stems from the fact that it facilitates predicting the consequences of interventions, since except for the elimination of influences upon the targeted variable the causal structure persists as before. Thus, the persisting causal structure can be used to trace out the intervention’s consequences. It is important for the purposes of this essay that modularity is explicitly incorporated in the definition of an ideal intervention, since non-ideal interventions that fail this condition will figure prominently in the subsequent discussion.

There are several features of Woodward’s theory that should be appealing to one who wishes to defend causal explanation in the social sciences. According to Woodward’s theory, laws of nature in the style of those of physics are not required for
causal explanation (cf. Woodward 2000). Another appealing feature of Woodward’s theory is that it provides a straightforward account of why causal explanations should be of interest in social science. A fundamental aim of social science is to provide a basis for rational policy making, which requires being able to predict the consequences of various sorts of interventions. Hence, it is of obvious interest to distinguish between generalizations that are invariant under intervention and those that are not, and to accord a special virtue to the former. Finally, Woodward’s theory is straightforwardly applicable to social science examples (cf. 1999). Indeed, the concept of invariance is inspired by the notion of a structural equation in econometrics—a set of equations is said to be structural when it can be used to predict the consequences of policy interventions.

Let us consider, then, what conclusions ensue regarding methodological individualism if Woodward’s theory is accepted. The most immediate implication is a confirmation of the argument that individual level mechanisms are not necessary for explanation in social science. There is no reason in principle to deny that a generalization relating two aggregate social features may be invariant under some range of interventions. Moreover, there are well-documented examples of such generalizations (e.g. concerning interest rates and inflation) in macro-economics (cf. Hoover 2001, 124-125). Hence, such a generalization may be used for causal explanation even if individual level mechanisms are neither explicitly described nor implicitly suggested.

However, there is more to be said about the matter than this. According to Woodward’s theory, there is both a threshold and a continuum with regard to invariance (2000, 214; 2003, 257). A generalization that is invariant under some interventions passes the threshold, and is therefore a potential basis for some explanations, but some
generalizations are invariant under a greater range of interventions than others. For example, consider the relation between the ideal gas law and statistical mechanics. Although both are invariant under some interventions, the scope of invariance of statistical mechanics is far greater than that of the ideal gas law. In addition, it is commonly thought that explanations on the basis of statistical mechanics are deeper. So, among those generalizations that pass the threshold, there is a continuum from more to less invariant that corresponds to a continuum from deeper to more shallow explanation. Examples of this sort are not limited to physics. For instance, the laws of Mendelian genetics are invariant under some interventions, yet their range of invariance is more narrowly restricted than generalizations of molecular genetics. These examples lend some support to the idea that generalizations concerning micro-mechanisms are typically invariant under a broader range of interventions than those relating aggregate features.

The methodological individualist could be interpreted as maintaining that this pattern holds true of social phenomena as well. For example, Mill argued that causal relations among collective social properties depend upon behavioral tendencies of individuals that often vary from one nation to the next (1856, 537-538, 570-573). Hence, Mill maintained that attention to the mechanisms by which these tendencies are formed was essential for understanding when collective level generalizations would and would not obtain. A similar idea is also mentioned by Woodward (2000) with regard to economics:

A standard assumption among many microeconomists—one might take it to be constitutive of a certain sort of methodological individualism—is that generalizations that will be invariant under such changes in information and
Methodological Individualism, Explanation, and Invariance

prices all describe the behavior of individual economic agents rather than the relations between macroeconomic or aggregate-level variables like ‘inflation’, ‘unemployment’, and ‘gross domestic product’. That is, the idea is that there are no purely macroeconomic relationships that are invariant under changes in information and incentives and hence that there are no fundamental explanatory relationships between macroeconomic variables. (P. 220)

As Woodward suggests, this conclusion could be regarded as a formulation of the methodological individualist position. In the remainder of this essay, I explicate and assess the merits of this version of methodological individualism.

Structure Altering Interventions

Assessing the claim that, in social science, generalizations concerning micro-mechanisms are invariant under a broader range of interventions than collective level generalizations requires some clarification of the notion of a “greater range of invariance.” The range of invariance of some generalizations is limited because they provide only an approximate, inexact representation of the causal relationship among the variables in question. But since one could inaccurately model a mechanism as well as a collective level relationship, such inexactitudes cannot be the basis of the individualist’s argument. Rather, the point must be that even if one has the collective level causal relationships right, they still will be invariant under a narrower range of interventions than individual level mechanisms. However, if the interventions in question are ideal, it is unclear how this could be so. For an ideal intervention, by definition, does not alter

---

3 The argument also resembles Little’s suggestion that methodological individualism might be motivated by the premise that regularities concerning individuals are “stronger” than those concerning collective social features (1991, 189).
causal influences emanating from the variable it targets. Hence, any correct representation of a causal relationship is invariant under any ideal intervention that targets the cause. For example, in figure 2, an ideal intervention on $S$ will not alter the influence of $S$ upon $L$. What this implies is that some notion of intervention other than ideal is presumed in the claim that micro-mechanisms are more robustly invariant than macro-level relationships. I suggest that the pertinent concept is what I call a \textit{structure altering intervention}.

Structure altering interventions are best understood through contrast with ideal interventions. Recall that an ideal intervention eliminates other influences upon its target, but does not otherwise alter the causal structure. For example, in figure 2(a), the intervention on $S$ does not alter the causal effect of $S$ upon $L$. Similarly, in 2(b), the intervention on $B$ does not alter the causal relationship between $A$ and $S$. That an ideal intervention does not alter the causal structure except by eliminating influences upon the targeted variable is the modularity condition embodied in item (i) of the definition provided above. A structure altering intervention, in contrast, makes changes in causal structure besides blocking the usual influences upon the targeted variable. Thus, structure altering interventions are \textit{not} modular.

The concept of a structure altering intervention can be defined as follows. Suppose the variable $X$ is a member of a set of variables $V$ that represent features of the system of interest. Then an intervention on $X$ is \textit{structure altering with respect to $V$} just in case it changes causal relationships among the variables of $V$ \textit{in addition to} eliminating the causes of $X$. For instance, in figure 2(b), $V = \{B, A, S\}$ and the intervention is on $B$. 

So, this intervention would be structure altering if it modified the relationship between \( A \) and \( S \). Thus, no structure altering intervention is an ideal intervention.

It is important to distinguish structure altering interventions from the difficulty of differentiating causation from mere correlation. As figure 2 illustrates, a correlation between two variables might result from one being a cause of the other or from the presence of a common cause of both. Thus, the correlation between \( B \) and \( S \) in figure 2(b) is a mere correlation: it is not invariant under any ideal intervention that targets \( B \).

In the barometer example, it is obvious that the correlation is due to a common cause, but distinguishing cause from mere correlation is often difficult, especially in social science. However, that problem is distinct from the possibility that a genuinely causal relationship might breakdown under certain types of intervention. For instance, the ideal gas law is invariant under some ideal interventions, and hence is causal according to Woodward’s theory. Nevertheless, the ideal gas law breaks down under interventions that set the pressure to a very high value. In this case, the fragility of the generalization is due to its rough and approximate character. But even a generalization that precisely describes a causal relationship in a given context might be subject to structure altering interventions. Causal relationships often depend upon background conditions too numerous and complex to fully and explicitly incorporate into a model. Consequently, interventions that change such background conditions may be structure altering. Even if the generalization accurately described the causal relationship under the original set of background conditions, it might be an inaccurate representation of that relationship in the new circumstances brought about by the intervention.
Consider James Scott’s account of the effects on the social structure of a Malaysian village of a government sponsored irrigation project that made it possible to grow two rice crops per year rather than just one (1985, 74-85). Initially, the project significantly improved the economic situation of villagers, from the land-poor who relied on wage labor to the larger land owners. The increase in the supply of rice sharply reduced the threat of famine and, initially, the additional rice production doubled the demand for field labor and thereby significantly raised the income of poorer villagers. However, land owners soon discovered that renting combine-harvester machines was better than hiring field labor under the new system, since double-cropping required quickly harvesting one crop so that the next could be planted. Not only did this turn of events adversely affect the wages of land poor villagers, it also undermined traditional demonstrations of generosity through which wealthy farmers attempted to ensure reliable sources of labor in the future. These practices included sumptuous feasts to which all in the village were invited, bonuses for laborers at the end of the harvest, and tenancy agreements that made allowances for poor harvests. In short, the innovation of double-cropping fundamentally altered the economic structure of mutual dependence between poor and wealthier villagers and all of the practices that went along with it.

Let $R$ indicate the annual rice production, $D$ the demand for field labor, and $E$ the wage earnings of villagers. Finally, let I represent the intervention, that is, the government sponsored double-cropping program. Then this example can be represented as follows:
The intervention on $R$ is structure altering with respect to \{R, D, E\}. The intervention significantly attenuated the influence of rice production upon demand for field labor, which is represented in 3(b) by the deletion of the arrow from $R$ to $D$. An arrow directly from the intervention to $D$ is included in the graph in 3(a) because the intervention affected the demand for labor through a path not passing through $R$. That is, growing two crops per year rather than one placed a higher premium on a quick crop harvest, thereby increasing the use of combine-harvester machines and hence undermining the demand for field labor. Notice that the causal relationship in 3(a) would be invariant under many interventions on $R$ that did not increase the number of crops per year. But the graph in 3(a) implicitly treats the harvesting of one crop per year as a stable background condition, and hence no longer correctly represents the influence of rice production upon the demand for field labor when that background condition is changed.

This example suggests that structure altering interventions directly affect more than one variable, and hence violate item (ii) in the definition of an ideal intervention. If that were always true, then the modularity clause in item (i) of that definition (that the intervention makes no changes in causal structure aside from eliminating other causes of the targeted variable) would be redundant. However, given the importance of modularity for our purposes, it is worthwhile to have it explicitly stated in the definition. Moreover,
an intervention might directly affect more than one variable without being structure altering.

Given the concept of a structure altering intervention, the question of whether social mechanisms generally have a significantly broader range of invariance than collective level relationships can be rephrased in the following way. Since an intervention is structure altering only with respect to a particular set of variables, an intervention might be structure altering with regard to one set of variables but not another. Let $S$ be the social system of interest, let $C$ be a set of variables representing collective features of $S$. Let $I$ be a set of variables representing those properties of individuals comprising $S$ that are pertinent to the collective features characterized by the variables in $C$. Then the claim at issue is the following:

\[(MI) \text{ For any social system } S, \text{ the set of interventions that are structure altering with respect to } I \text{ is significantly smaller than that for } C.\]

The Malaysian example illustrates the intuition underlying (MI). In that example, double-cropping undermined the positive effect rice yields exerted upon the demand for field labor. Hence, that intervention was structure altering with respect to the variables $\{R, D, E\}$. Yet the process is easily understandable in terms of a model that focuses on the decisions of individual cultivators. In such a model, the quantity of field labor demanded by a given individual might be a function of his expected crop yield as well as the relative costs and efficiencies of field labor and the mechanical harvesters. The irrigation project that made double-cropping possible simply changed the values of some these variables but not their relation to the demand for field labor. Consequently, the defender of methodological individualism could argue that the intervention would not be
Methodological Individualism, Explanation, and Invariance

structure altering with respect to the set of variables that would be included in a good rational choice model of the situation. The issue, then, is whether there is some reason to think that (MI) is true in general, or whether it is merely an accidental feature of the example just described.

The Multiple-Realizability Argument

The multiple-realizability argument is possibly the most common objection to methodological individualism (cf. Little 1991, 190-195; Kincaid 1996, 145-155; Sawyer 2002; Zahle 2003) and reductionism generally (cf. Kitcher 1984). The central thesis of the multiple-realizability argument is that a single macro-level generalization is sometimes instantiated by several micro-level mechanisms. In such circumstances, the argument concludes, the macro-level relationship is more unified and hence a better explanation. Thus, the holist might claim, when a macro-level relationship is multiply realized it will be invariant under a broader range of circumstances than any particular micro-mechanism.

Whatever the merits of the multiple realizability argument in general, I think that this application of it is mistaken. To see why, look again at (MI). This proposition requires that, for each system, the set of interventions that are structure altering with respect to micro-level variables is smaller than the set that is structure altering with regard to the collective level variables. This does not entail that either the configuration of the mechanism components or the relation of micro-mechanism to collective is the same across all systems of interest. For instance, although the Mendelian gene is multiply realized at the level of molecular genetics, in any particular case causal
relationships at the level of molecular genetics are more robustly invariant than those at
the level of Mendelian genetics. In short, a generalization whose scope is very limited
may nevertheless be extremely robust under intervention, while a generalization with an
extremely broad range of application may be very fragile. Another way to put the point
is by distinguishing between the following two questions:

1. At which level does one find the simplest, most broadly recurring patterns?
2. At which level does one find relationships least prone to structure altering
   intervention?

These are two different questions, and the level given in response to (1) may differ from
that for (2). For abstracting from micro-detail often allows for the identification of
simple patterns with a broad scope of applicability. Yet precisely this abstraction from
micro-detail yields generalizations that omit relevant factors and consequently which are
more susceptible to structure altering interventions.\(^4\) The methodological individualist,
then, could assert that this is true of social science. Although collective social
relationships may be multiply realized by individual level mechanisms, the individualist
could argue that micro-mechanisms are, invariant under a broader range of interventions
for any given social system.

A different holist critique of (MI) would be to maintain that collective level
models can be made less susceptible to structure altering interventions through the
addition of further variables. For instance, in the Malaysian example, the intervention is
structure altering with respect to \(\{R, D, E\}\); but suppose we added the variables \(C\) and \(H\),
which represent the number of rice crops per year and the frequency of use of combine-

\(^4\) See (Steel 2004, van Bouwel 2004) for further elaboration of the case that the multiple realizability
argument is an objection to certain versions of reductionism but not others.
harvester machines, respectively. It is plausible that the intervention would not be structure altering with regard to \{R, D, E, C, H\}. That is, in a model that accurately represented the causal relationships in the Malaysian example, the effect of \(R\) upon \(D\) would depend upon \(H\) which in turn depends upon \(C\). The causal relationships specified by such a model would not be altered by the intervention. Rather, the intervention would raise \(C\) and consequently \(H\) and thereby reduce the effect of \(R\) upon \(D\), just as the model describes.

In general, structure altering interventions change the background conditions upon which the causal relationships in question depend. A critic of methodological individualism, then, could maintain that this relevant background condition can be represented by a macro-level variable, and that the intervention might not be structure altering with regard to a set of macro-level variables that included this additional one. If this strategy were generally successful, then there would seem to be little basis for the claim that individual level mechanisms are invariant under a significantly broader range of interventions that aggregate relationships.

The best individualist response here, I think, is that without an understanding of underlying mechanisms such variables would be added in a largely ad hoc manner. By analogy, imagine someone who wished to argue that statistical mechanics was in fact no more fundamental than the ideal gas law. This person might maintain that for any intervention under which the ideal gas law breaks down, some further condition stated in terms of higher level properties can be found to patch it up. Even if this were true, it would hardly constitute a strong argument against the claim that statistical mechanics provides deeper explanations. For without statistical mechanics, one would have little
Methodological Individualism, Explanation, and Invariance

understanding of why *those* conditions mattered and one would be unable to anticipate new circumstances in which the ideal gas law fails. Thus, the methodological individualist argument would be that the relationship between collective level generalizations and micro-mechanisms in social science is similar in this regard to that of phenomenological and fundamental laws in physics.

For instance, consider Jeffrey Paige’s (1975), *Agrarian Revolution*, which proposes a theory about which types of agrarian systems lead to which types of conflict between cultivators (e.g. peasants) and non-cultivators (e.g. land owners). Kincaid (1996, 70-80) cites Paige’s work as an example of quality social science done from a holist perspective. And indeed, Paige proposes general hypotheses linking particular types of agrarian social structure to particular types of conflict. For example, Paige proposes that agrarian revolutions are most common in systems in which the income of non-cultivators is based upon control of land and in which cultivators are paid wages, either as migratory laborers or through a share cropping system (1975, 58-70). In contrast, when the wealth of non-cultivators stems from capital and cultivators are paid in wages, reform movements in which workers form unions and endeavor to win concessions (e.g. higher pay) are more likely (1975, 48-58). Kincaid uses Paige’s study to support his case against individualism primarily by arguing that many of the terms in Paige’s explanation are multiply realized at the individual level (1996, 163-164). As we have seen, multiple-realizability poses no real threat to (MI). But Kincaid also uses Paige’s example to argue that individual level mechanisms are not necessary for refining macro-level generalizations.
Methodological Individualism, Explanation, and Invariance

We can refine and improve macrosociological explanations in macrosociological terms by bringing in other macrosociological variables. Paige, for example, did exactly this when he added in the effect of urban political parties to improve and test his basic model. … So there is no reasonable argument that micromechanisms, in particular individual mechanisms, are necessary for predictive improvement. (P. 181; italics in original)

However, an examination of *Agrarian Revolution* reveals that Paige’s general strategy for dealing with anomalous cases relies essentially upon individual level mechanisms.

Paige’s strategy is to argue that anomalous cases are in fact consistent with the individual level mechanism he proposes to explain his postulated macro-relationships. For instance, consider how Paige treats an important exception to his hypothesis that non-cultivator’s wealth based upon capital and cultivator’s upon wages leads to reforms rather than revolution (1975, 50-58). The pertinent feature of the mechanism in this case is that the response of non-cultivators to demands of cultivators (e.g. for higher wages) depends upon the ability of the non-cultivators to increase profit margins through technological investments that raise efficiency. This mechanism explains the macro-level regularity insofar as non-cultivators whose wealth derives from industrial capital, but not those whose wealth stems from land ownership, are typically capable of making such investments. In this context, exceptions to the rule that a capital/wages agrarian system yields reform movements rather than revolutions are accounted for by showing that the exceptional cases are ones in which the non-cultivators were unable to increase efficiency through technological investments. Thus, Paige’s explanation of the exception is not ad hoc precisely because it shows that the anomalous case in fact follows the pattern of the
Methodological Individualism, Explanation, and Invariance

proposed underlying mechanism. A very different assessment of Paige’s hypothesis would be in order if he had, for each anomalous case, simply pointed to some collective level feature that allegedly explained the problem away.

Are Social Mechanisms Robustly Invariant?

The version of methodological individualism described here might be false even if it is not undermined by the multiple-realizability argument. Thus, it is important to consider whether there is any general argument for the premise that individual level mechanisms are invariant under a broader range of interventions than collective level generalizations. The quotation from Woodward given above suggests an argument along the following lines. Attention is restricted to interventions that change the information or incentives of agents. Then the claim is that an accurate rational choice model would be invariant under any such intervention. That is because such models presume that any changes to information or incentives only affect agents’ beliefs or the consequences associated with particular actions. Meanwhile, agents are assumed to possess stable preferences that are independent of the decision task with which the agent is confronted. Given that one can model changes in incentive structure and the manner in which agents accommodate new information, agents’ choices can be derived from their preference rankings. Thus, the argument concludes, an accurate rational choice model will be invariant under any intervention that targets information or incentives, whereas the same is not true of macro-sociological generalizations.

This argument treats preferences as relatively stable causes of decisions rather than as things shaped by the decision making process itself. Yet several psychologists
have maintained that preferences are, to a large extent, constructed in the process of making decisions (Slovic 1995; Simon et al. 2004). Experimental results concerning “preference reversals” are among the most striking lines of evidence for this view of preferences (cf. Hausman 1992, chapter 13; Slovic 1995; Guala 2005, 91-101). Preference reversals call into question “procedure invariance,” which “requires strategically equivalent methods of elicitation to yield the same preference order” (Tversky et al. 1990, 204). If procedure invariance is false, then distinct framings of strategically equivalent decisions may result in the formation of distinct preferences.

Classic preference reversal experiments offer subjects a choice between two bets, one that offers a small payoff with high probability (the P-bet) and another that offers a large payoff with low probability (the $-bet) (cf. Slovic 1995, 365-366). After subjects choose their bet, they are asked to specify a price for each. Since stated prices might reflect strategic market considerations rather than the value of the item to the person, the prices are elicited in a context in which strategic pricing is not possible. The most commonly used is the Becker-DeGroot-Marschak (BDM) elicitation method (cf. Roth 1995, 19-20). In this method the subject is asked to state a price for the bet on the conditions that a buying offer is generated randomly and must be accepted if it meets the subject’s stated price, while the subject plays the bet immediately otherwise. Since the offer is randomly generated, it cannot be affected by the posted price. Moreover, the subject is not able to search for other buyers if the initial offer is not to her liking. Hence, the pricing strategies that might be pursued by a real world merchant cannot be used in

---

5 Procedure invariance is also sometimes designated by the expression “context-free preferences” (cf. Cubitt et al. 2004, 709).
Methodological Individualism, Explanation, and Invariance

this context, and the dominant strategy for the subject (if she is a utility maximizer) is to state the price that reflects the value of the bet to her.

The surprising result is that, under these circumstances, a substantial proportion of the subjects choose the P-bet but assign a higher price to the $-bet. This situation is known as the “expected” preference reversal. Given the natural assumption that people prefer more money to less, such experiments seem to show that an agent’s preferences, far from being context independent, depend on how a choice or task is framed. Some economists have attempted to make preference reversals go away in modified versions of the experiment (cf. Grether and Plott 1979). Others have endeavored to show that the preference reversals are not due to a failure of procedure invariance but rather can be pinned on the BDM elicitation method (cf. Holt 1986; Segal 1988). However, preference reversals have proven an extremely persistent phenomenon, and such attempts to explain them as experimental artifacts have not been successful (cf. Tversky et al. 1990; Camerer 1995, 658-665; Cubitt et al. 2004; Guala 2005, 121-128).

But although preference reversals are a real laboratory phenomenon, how widespread they are outside the laboratory walls is still debatable (Guala 2005, 225). I found two distinct explanations of preference reversals that have strikingly different implications about this matter. The first explanation is that preference reversals are due to scale compatibility (Tversky, Slovic, and Kahneman 1990, 211). Scale compatibility is the hypothesis that “the weight of any aspect (for example, probability, payoff) of an object of evaluation is enhanced by compatibility with the response (for example, choice, pricing)” (ibid). Thus, when choosing between the P-bet and the $-bet, subjects focus on the probability and accordingly take the P-bet. But when asked to provide a monetary
Methodological Individualism, Explanation, and Invariance

valuation, subjects focus on the monetary scale, and hence rank the $-bet more highly. If this explanation is correct, then preference reversals certainly cannot be dismissed as some oddity of the laboratory with no relevance to real life. Preferences are expressed in various ways, of which overt choices and prices are merely two common examples. So, if preference orderings are indeed so closely linked to the scale on which the options must be ranked, then preference reversals should be a prevalent feature of everyday life.

But scale compatibility is not the only possible explanation of preference reversals. Consider a person in a preference reversal experiment. If the person is risk averse, then she will choose the P-bet when given the choice of playing one or the other. But now consider the question: which bet has a higher monetary value? The subject might naturally interpret this question as a request for an estimate of the market value of each bet. Indeed, outside the laboratory, questions about prices normally are requests for information about market rather than personal value. For concreteness, suppose that the P-bet gives a 99% chance of winning 10 dollars, while the $-bet gives a 0.099% chance of winning 10,000 dollars. While it is obvious that no one will pay more than $10 for the P-bet, it is likely that some risk-seeking individuals will agree to pay more than $10 for the $-bet. Hence, in a real market, a seller can expect to obtain a higher price for the $-bet than for the P-bet. Consequently, it is understandable that the subject might assign a higher price to the $-bet, even if she would play the P-bet when given a choice between the two.

The possibility that this simple insight might explain preference reversals is developed in a paper by Xiaoyong Chai (2005). Let us use the expression “market price reversal” to refer to the hypothesis that subjects in preference reversal experiments often
interpret questions about the prices of bets as questions about their market value. Market price reversal can explain a puzzling aspect of the data found from some of the earliest experiments. It turns out that expected preference reversals are far more common when subjects are asked to specify a selling price rather than a buying price (cf. Seidl 2002, 622-3), a result that Chai replicates (2005, 190). A question about the selling price is very likely to be interpreted as a question about market value, whereas a question about buying price is more likely to suggest an assessment of personal value to the subject. Moreover, as Chai observes, a prospective buyer of the $-bet might hope to find a risk averse person who possesses it and who is willing to sell it for less than its expected value (2005, 182). Consequently, market preference reversal predicts that “unexpected” preference reversals (in which the subject chooses the $-bet but prices the P-bet higher) should be more common when a buying rather than a selling price is asked for, which is indeed correct (Seidl 2002, 623; Chai 2005, 191). In contrast, the scale compatibility hypothesis does not explain why the frequency of expected and unexpected preference reversals is linked to whether a selling or buying price is requested. For posing a question about price in terms of selling rather than buying does not change the scale on which the valuation is made.

The primary objection to the market price reversal hypothesis is that elicitation procedures like BDM are designed precisely for the purpose of eliminating market considerations. And indeed, the rate of expected preference reversals is significantly less in the BDM elicitation method than when subjects are merely asked to specify a selling price (cf. Seidl 2002, 623). However, there are several reasons why BDM and other such elicitation methods might not eliminate market pricing effects. The first is simply that
many subjects may not appreciate the strategic implications of the BDM procedure. The subject is required to quickly learn the new rules of the game played in the experiment and may not have time to carefully devise an optimal strategy in response to them. Under such circumstances, a subject might follow a heuristic that works reasonably well in a real world context that appears similar to the experimental situation. Thus, many subjects may still propose selling prices that are estimates of market values. Moreover, as Chai observes, in the BDM procedure “the absence of a human buyer is unverifiable to the subjects” (2005, 185). That means that the subject cannot directly verify that her asking price has no affect on the offer. In light of this, Chai modifies the BDM mechanism so that the offer is present as physical money in a sealed envelope placed before the subject (2005, 186-187). The subject, therefore, can easily see that the price she posts has no effect on what is offered. Chai found that the difference in relative frequency between expected and unexpected preference reversals was not statistically significant when the envelope method was used (2005, 191).

Thus, the market price reversal hypothesis attributes preference reversals to a systematic difference between real markets and experimental circumstances. In real markets it is normally the case that (a) sellers are free to search out buyers (and vice versa) and (b) the posted selling price is not independent of the offer. In the context of the experiment using the BDM elicitation method, neither (a) nor (b) obtains. Nevertheless, questions about the prices of lotteries are likely to cue the application of reasoning heuristics appropriate to real markets. According the market price reversal hypothesis, then, preference reversals in experiments can in fact be regarded as violation of procedure invariance, since subjects do give conflicting rankings to strategically
Methodological Individualism, Explanation, and Invariance

equivalent options. But that result does not entail that violations of procedure invariance
are widespread in economic or other social contexts. For if market price reversal is the
correct explanation, these experiments show only that it is possible to trick people with
decision scenarios that strategically differ from the situations in ordinary life that they
superficially resemble.

Which of these two explanations of preference reversals is correct (if either)
remains an open question. What they illustrate is our present state of ignorance
concerning the manner in which preferences are or are not shaped by specific decision
contexts. Moreover, preference reversals cast doubt on the argument described in this
section for the claim that individual level mechanisms are more robustly invariant than
collective level generalizations. Indeed, given the importance of decision making to
individual level mechanisms, any argument that they are robustly invariant must rest on
some premise about the nature of preferences. An assessment of the merits of
methodological individualism as construed here, then, must await further advances in our
understanding of preference formation.

Conclusion

Disputes concerning methodological individualism depend upon premises about
what makes some explanations deeper than others. Consequently, methodological
individualism can be fruitfully addressed only within the context of some reasonably
clear and well motivated theory of explanation that is capable of providing answers to
such questions. I examined methodological individualism from the perspective of
Woodward’s theory of causal explanation, according to which invariance under
intervention is the distinguishing feature of causal generalizations. This theory supports
the claim that individual level mechanisms are not always required for explanation, but it
also suggests an argument for the conclusion that such mechanisms are needed for deep
explanation. Moreover, multiple-realizability does not pose a genuine challenge to
methodological individualism so construed. Nevertheless, the positive case for the robust
invariance of individual level mechanisms is called into question by experimental results
concerning preference reversals. Indeed, any argument attempting to show that micro-
mechanisms are more robustly invariant than collective level generalizations will rest on
premises concerning the nature of preferences and their formation, matters which remain
quite uncertain. Consequently, the adequacy of methodological individualism as
construed here depends on unresolved empirical questions.

References

587-703.

177-194.

Economic Journal 114: 709-726.

Press.
Methodological Individualism, Explanation, and Invariance


Methodological Individualism, Explanation, and Invariance


Methodological Individualism, Explanation, and Invariance


Methodological Individualism, Explanation, and Invariance


_Daniel Steel is an assistant professor in the Department of Philosophy at Michigan State University. His research interests lie primarily in issues of causation as they pertain to social science and biology._