This is a preprint of a paper published in the American Political Science Review 107(4): 629-643 (2013).

Methodological Individualism and Holism in Political Science: A Reconciliation

Christian List and Kai Spiekermann

London School of Economics and Political Science*

Abstract: Political science is divided between methodological individualists, who seek to explain political phenomena by reference to individuals and their interactions, and holists (or non-reductionists), who consider some higher-level social entities or properties such as states, institutions, or cultures ontologically or causally significant. We propose a reconciliation between these two perspectives, building on related work in philosophy. After laying out a taxonomy of different variants of each view, we observe that (i) although political phenomena result from underlying individual attitudes and behaviour, individual-level descriptions do not always capture all explanatorily salient properties, and (ii) non-reductionistic explanations are mandated when social regularities are robust to changes in their individual-level realization. We characterize the dividing line between phenomena requiring non-reductionistic explanation and phenomena permitting individualistic explanation and give examples from the study of ethnic conflicts, social-network theory, and international-relations theory.

---

* Christian List is Professor of Political Science and Philosophy, Departments of Government and Philosophy, LSE, London WC2A 2AE, U.K. Kai Spiekermann is Associate Professor of Political Philosophy, Department of Government, LSE, London WC2A 2AE, U.K. An earlier version of this paper was presented at the LSE Workshop on "Reductionism and Non-Reductionism in the Social Sciences", June 2012. We thank the participants at this event, Michael Baurmann, Peter Menzies, Philip Pettit, four anonymous reviewers, and the editors for helpful comments or discussions.

Introduction

Political science is divided between those who think that a scientific approach to the study of politics requires methodological individualism and those who consider this idea hopelessly reductionistic. Crudely put, *methodological individualism* is the thesis that good social-scientific explanations should refer solely to facts about individuals and their interactions, not to any higher-level social entities, properties, or causes. *Holism*, or more neutrally *non-reductionism*, is the negation of this thesis. Political scientists influenced by economic methodology, especially rational choice theorists, are usually methodological individualists, while more sociologically or historically minded scholars and many institutionalists, international-relations scholars, and normative political theorists lean towards non-reductionism. A key point of contention is the status of collective entities such as states, nations, ethnic groups, cultures, political parties, and other institutions. Are these mere by-products of individual behaviour, or are they of independent ontological and/or causal significance?

We propose a reconciliation between these two perspectives, and show that there are insights, but also mistakes, on both sides. To make methodological progress, we must recognize these insights and overcome the mistakes. A reconciliation becomes possible once we see that there is not just one version of each view, but many, and that being an individualist in some respects is compatible with being a holist in others. Our proposal builds on ideas from the philosophy of mind (e.g., Putnam 1967; Fodor 1974; Stoljar 2009, 2010; Kim 1998, 2005; List and Menzies 2009) and the philosophy of social science (e.g., Macdonald and Pettit 1981; Kincaid 1986; Pettit 1993, 2003; Tollefsen 2002; Sawyer 2002, 2003; List and Pettit 2011; Greve 2012), as well as the existing

debate in political science (e.g., Satz and Ferejohn 1994; Wendt 1999; Hay 2006; Peters 2012, ch. 1).

Methodological individualists are right to remind us that the social world is ultimately the *result* of many individuals interacting with one another and that any theory that fails to accept this basic premise rests on mysterious metaphysical assumptions. We call this premise "supervenience individualism". "Supervenience" is a philosophical term of art that refers to the necessary determination of one set of facts by another. Facts about molecules or chemical compounds, for example, *supervene on* (are determined by) facts about the atoms and particles of which they consist. Similarly, facts about social phenomena *supervene on* facts about individuals and their interactions. From this, however, it does not follow that all social phenomena can be *explained* in individualistic terms alone. This mirrors the observation that, despite the supervenience of chemical and biological phenomena on microphysical ones, chemical and biological *explanations* cannot necessarily be reduced to purely physical ones. In a slogan, "supervenience" does not imply "explanatory reducibility" (as observed by Kincaid 1986 and Sawyer 2002, 2003).

Holists are therefore right to insist that some social-scientific explanations must employ non-individualistic terms. It is entirely possible that *social* properties, not *individual-level* ones, display the most systematic causal relations in some social phenomena. The relation between interest rates and inflation, for example, may well be more robust than any particular individual-level transmission mechanism. This *explanatory* holism, however, does not imply any more radical form of holism that gives some kind of *metaphysical* priority to social structures over and above the individuals living in them. Again in a slogan, "explanatory holism" does not imply "radical social-

fact holism". Given how fragmented political science is, there is much confusion on these issues and little agreement on how individualistic or holistic the discipline should be. The proposed reconciliation is intended to move the debate beyond its current stalemate.

The closest precursor to our analysis can be found in Sawyer's (2002, 2003, 2005) defence of "non-reductive individualism" in the social sciences, an approach that "accepts that [fundamentally] only individuals exist but rejects methodological individualism" (2002, p. 537). Sawyer was one of the first to embark on a comprehensive translation of relevant ideas from the philosophy of mind into the philosophy of social science. Here, we address the methodological debate in political science, which has not yet taken on board the insights from related philosophical debates. We develop a refined taxonomy of different forms of individualism and holism and look at where political scientists appeal to them; and we offer a novel characterization of phenomena that require non-reductionistic explanations, drawing on recent work in the study of "mental causation", especially on higher-level causal relations that are robust to changes in their microrealization (List and Menzies 2009).

We begin with a sketch of the history of methodological individualism and holism, followed by our taxonomy of different variants of each view. We then present our argument for the compatibility of "supervenience individualism" and "causal-explanatory holism". This allows us to reassess the debate in political science and to identify conditions characterizing systems or phenomena that call for holistic as opposed to individualistic explanations. To illustrate these conditions, we briefly consider the study of ethnic conflicts and civil war (e.g., Fearon and Laitin 2003), social-network theory (e.g., Christakis and Fowler 2009), and the debate about the right level of explanation in international-relations theory (e.g., Wendt 2004).

A brief historical sketch

Individualism

The idea that the individual should be the ultimate unit of social analysis goes back at least to social-contract theorists such as Hobbes and Locke and featured in J. S. Mill's claim that the laws of society must be derived from the laws of individual psychology (Mill 1974 [1843]; Udehn 2001, pp. 43-49). Weber famously criticized the use of collective notions such as "*Volk*", "class", and "society" in social explanations and argued that we can achieve understanding ("*Verstehen*") only by interpreting the intentional actions of individuals:[1]

> "[C]ollectivities must be treated as *solely* the resultants and modes of organization of the particular acts of individual persons, since these alone can be treated as agents in a course of subjectively understandable action." (Weber 1978 [1922], p. 13, emphasis in the original)

This view, which Schumpeter (1908) called "methodological individualism",[2] was echoed by Hayek and Popper, who held that "abstractions" such as "society", "economic system", and "capitalism" should be avoided in social explanations (Hayek 1942, p. 286;

---

[1] For Weber, only individuals could be intentional agents. If collectives could be agents (as argued, e.g., in List and Pettit 2011), he might accept them as basic units of analysis.

[2] He did not defend it as a general methodological principle, but advocated methodological pluralism, taking the right unit of analysis to be guided by the research question.

Popper 1944, 1945a, 1945b).[3] On this view, "explanations of large-scale social phenomena (say, inflation) in terms of other large-scale phenomena (say, full employment)" were at best "half-way" explanations but never "rock-bottom" ones (Watkins 1957, p. 106; cf. Heath 2011; Udehn 2001, ch. 7; Lukes 1968):

> "[W]e shall not have arrived at rock-bottom explanations of such large-scale phenomena until we have deduced an account of them from statements about the dispositions, beliefs, resources, and inter-relations of individuals." (Watkins 1957, p. 106)

More recently, Elster (1985) defended "the doctrine that all social phenomena … are in principle explicable in ways that only involve individuals – their properties, their goals, their beliefs and their actions" (p. 5). He rejects Marxist functionalist tendencies to make claims such as "the capitalist class keeps unemployment high to suppress the working class", which ascribe intentions to collectives without recognizing that these are not purposive, intentional actors, and that even if they were, this would require further explanation (Heath 2011). Individualistic explanations, Elster (1985, p. 5) argues, "reduce the time-span between explanans and explanandum" and thereby the risk of mistaking spurious correlations for explanations. Ontologically, "the elementary unit of social life is the individual human action" (Elster 1989, p. 13). As can be seen, methodological individualism has been endorsed from a combination of ontological and explanatory perspectives (for recent reviews, see Udehn 2002 and Hodgson 2007).

---

[3] For Hayek, a methodological individualist "refrains from treating these pseudo-entities as 'facts', and … systematically starts from the concepts which guide individuals in their actions and not from the results of their theorising about their actions" (1942, p. 286).

Holism

The term "holism" was coined by the controversial South African politician and part-time academic Jan Smuts in his 1926 book *Holism and Evolution*. The association with Smuts may have given the term a bad repute. He notoriously supported South Africa's racial segregation (but also co-authored the preamble of the UN Charter, corresponded with Einstein, and advocated humanitarian values abroad). Nowadays the term is often used by its critics as a label for implausible metaphysical claims. We nevertheless use "holism" to refer to the opposite of individualism, without any negative connotation intended.

Non-individualistic approaches to the social sciences can be traced back at least to Durkheim (1982 [1901]), who introduced the notion of a "social fact":

> "A social fact is any way of acting, whether fixed or not, capable of exerting over the individual an external constraint; or: which is general over the whole of a given society whilst having an existence of its own, independent of its individual manifestations." (p. 59)

Social facts, for Durkheim, go beyond their "individual manifestations":

> "Whenever elements of any kind combine, by virtue of this combination they give rise to new phenomena. One is therefore forced to conceive of these phenomena as residing, not in the elements, but in the entity formed by the union of these elements. The living cell contains nothing save chemical particles, just as society is made up of nothing except individuals. Yet it is very clearly impossible for the characteristic phenomena of life to reside in atoms of hydrogen, oxygen, carbon and nitrogen." (p. 39)

Similarly, Durkheim suggests, a society can have properties that are irreducible to the properties of individuals. This kind of holism about properties, however, is consistent with the claim that "society is made up of nothing except individuals"; we return to this observation later. How exactly individual agency and social constraints relate to each other is often called the "structure-agency" question. Parsons (e.g., 1937) sought a middle ground between structural and individualistic explanations, interpreting social structures as systems of norms that can serve as reasons for individual actions, but still considering social structures causally relevant.

In recent research, we find at least three motivations for non-individualistic explanations. First, there has been a "systemic" or "institutional turn" in several areas of political science and related fields. Properties of "systems" or "institutions" are seen by many not only as *explananda* (or dependent variables, which are to be explained) but also as *explanantia* (or independent variables, which play some explanatory role). Theorists of democracy, for example, have turned their attention to large-scale "deliberative democratic systems", subsuming "the norms, practices, and institutions of democracy" (Mansbridge et al. 2012) and spanning across a range of domains, from local to global (see also Dryzek 2010). "Social-ecological systems" are the units of analysis in recent social-scientific studies of sustainability (e.g., Ostrom 2009), and institutions are sometimes seen as the units of selection in social evolutionary processes, though it remains controversial whether such explanations can ultimately be reduced to an individual level.[4]

---

[4] How reductive or non-reductive evolutionary explanations should be remains contested in the philosophy of biology too. Although molecular genetics suggests a reduction base

A second motivation for holism comes from a re-awakened interest in social ontology (e.g., Bratman 1999; Gilbert 2006; Tuomela 2007, 2010; Pettit 2003; Tollefsen 2002; List and Pettit 2011). Although most scholars accept that phenomena such as collective intentions, social norms, conventions, institutions, and group agency are ultimately *brought about* by individual attitudes and actions, many doubt that these phenomena are fully *explicable* in terms of individual attitudes and actions alone. A statement such as "The Conservatives want to come across as a party that cares about the environment" need not be straightforwardly reducible to a set of statements about the attitudes of individual party members. The collective intention may exist even though not all, or a majority, of them have a matching individual intention. The party may be committed to certain views because of earlier commitments, irrespective of its members' attitudes right now (List and Pettit 2011). Similarly, our acceptance of printed slips of paper as money, with all the conventions that go along with this, may be hard to explain without reference to joint intentions or social ontology (e.g., Searle 1995, 2010).

A third motivation for holism is the problem of complexity (e.g., Jervis 1997). Holists maintain that a reduction of social explanations to the individual level, while perhaps *logically* possible, is not possible *in practice*, if this requires that "our current theories could be extended or replaced with well-developed ones that allowed for reduction" (Kincaid 1986, p. 494). Holists may grant that all social properties are ultimately determined by underlying individual properties. But this does not imply that a

---

for biological evolution (Dawkins 1986), the complex nature of evolutionary processes makes this reduction difficult (Kitcher 1984; Rosenberg 1997). A focus on organisms or systems rather than genes has recently re-gained prominence (e.g., Godfrey-Smith 2009).

reduction is feasible, especially because the same social property can be instantiated by many different constellations of individual properties: the problem of "multiple realizability" (Putnam 1967; Fodor 1974). The fact that "Japan has a high savings rate" may be explanatorily significant for certain phenomena, yet it may be realized in so many different ways that a reduction is not feasible (Little 1991, pp. 190-5). Think of all the different ways of holding money, the different social and legal practices related to it, its different possible distributions in the population, and so on. Many if not most social properties are multiply realizable. Similarly, complex social systems, such as the social networks in which political opinions form and spread, may have to be analysed in terms of aggregate properties. Attempts to understand them by reference to individuals alone, without appeal to their systemic properties, may be hopeless (see, e.g., Epstein 2006).

Varieties of individualism and holism

The key insight required for a reconciliation between individualism and holism is that one can consistently be an individualist in some respects while being a holist in others. Some of the most compelling individualist theses are entirely compatible with the most plausible holistic ones (as also observed by Sawyer 2002, 2003 in his defence of "non-reductive individualism"). The debate in political science often misses this point, by collapsing different variants of each view into a single doctrine, vacillating, for example, between ontological and explanatory claims. As Hay (2006) notes, much of the debate revolves around "two mutually exclusive positions … often defined in mutually

antagonistic terms" (p. 89), and "unacknowledged ontological choices underpin major theoretical disputes" (p. 79).[5]

We now present a taxonomy of different variants of individualism and holism, based on a structurally similar taxonomy from the philosophy of mind (Stoljar 2009, 2010). Philosophers of mind are interested in the relationship between physical and mental phenomena. The relevant physical phenomena are those that can be described by viewing organisms and brains as physical systems (e.g., neural states and brain-physiological processes), and the mental phenomena are those whose description requires psychological language (e.g., beliefs, desires, and intentions). Central questions are how

---

[5] He continues: "whilst such disagreements are likely to be manifest in epistemological and methodological choices, these are merely epiphenomena of more ultimate determinate ontological assumptions" (p. 79). Arrow (1994, p. 1) identifies an ontological-explanatory conflation in methodological individualism, noting that because "[individual] decisions interact to produce an outcome which determines the workings of the economy", "it seems commonly to be assumed that the individual decisions then form a complete set of explanatory variables". Similarly, Vermeule (2011, p. 9) draws attention to two common fallacies in social science: "if the components of an aggregate ... have a certain property, the aggregate ... must also have that property" and "if the aggregate has a certain property, the components ... must have the same property". Finally, "individualism" is often equated with a commitment to rational-choice explanations. Wedeen (2002, p. 717) writes: "insofar as individualism presupposes agents who are forward-looking strategists forever calculating costs and benefits, there will be a serious ontological and epistemological divide between most rational choice and interpretivist theorists".

someone's psychological states relate to his or her brain and bodily states, and how physical processes in the body give rise to such higher-level phenomena as consciousness and first-person experiences. If we substitute "individual" for "physical" and "social" for "psychological" or "mental", the parallels between these questions and ours become evident. Although these parallels have been recognized before (e.g., Pettit 1993; Sawyer 2002, 2003; Greve 2012), the relevant taxonomy of philosophical views and its logical structure are not as widely appreciated in the social sciences as they should be.[6]

Philosophers of mind distinguish between at least four different variants of "physicalism", each of which is a distinct thesis about how phenomena at the physical and psychological levels are related. They concern the relationships between *facts* at these two levels, between *particular objects* (e.g., entities, events, and processes, as explained below), between *general properties*, and between *causal explanations*. Each thesis has a social-scientific counterpart. Their negations correspond to different variants of holism.

---

[6] We follow and adapt Stoljar's taxonomy from the philosophy of mind. The key elements also occur in Sawyer's framework, though sometimes in subtly different and, for our purposes, less general forms. E.g., Sawyer considers the "token identity" thesis only with respect to events (under what he describes as a "particularistic theory with a commitment to properties") and takes it to imply supervenience individualism. We permit a more general ontology, in line with Stoljar's. Here, supervenience individualism is compatible with token holism, construed as a thesis about objects more generally (which may be defensible with regard to social entities such as Facebook, the Supreme Court, and so on).

A thesis about facts

Scientific theories usually refer to, and offer explanations of, certain domains of *facts*. As already anticipated, different theories employ different concepts and categories to *describe* the facts they are concerned with.[7] That a hydrogen atom contains only one proton and that human beings are capable of linguistic communication are examples of physical and psychological facts, respectively. The most basic physicalist thesis is:

> **Supervenience physicalism:** The physical facts fully determine the psychological facts; i.e., any possible worlds that are identical with respect to all physical facts will necessarily be identical with respect to all psychological facts.

Supervenience physicalism is the mark of most scientifically grounded worldviews: though *described* in non-physical terms, psychological facts are fully *determined* by more fundamental physical facts.[8] It is hard to imagine a world in which all physical particles, forces, and so on are configured in *exactly* the same way as they are in our world – including all the particles making up all living creatures and their environments – but where some psychological features are different. Supervenience physicalism becomes

---

[7] Heuristically, a fact is something whose description can be correctly substituted into the expression "It is the case that _". Formally, facts can be represented by propositions.

[8] Supervenience physicalism can be equivalently defined in terms of properties. But since we have not yet introduced the notions of objects and properties, we here give a definition in terms of facts. This is consistent, e.g., with the exposition in Chalmers (1996).

more or less demanding depending on how narrowly or broadly we demarcate the realm of facts that count as "physical".[9] The social counterpart, as already foreshadowed, is:

> **Supervenience individualism:** The individual-level facts fully determine the social facts; i.e., any possible worlds that are identical with respect to all individual-level facts will necessarily be identical with respect to all social facts.

Examples of individual-level facts are that the individuals in the US acted in certain ways on November 6, 2012, and that some went to the polling station. Examples of social facts are that Obama was re-elected President and that voting turnout decreases as the laws make voter registration harder, other things being equal. In analogy with supervenience physicalism, supervenience individualism can be more or less demanding, depending on what we take the individual-level facts to be. On a narrow definition, these may be confined to facts about the attitudes and behaviour of each individual in isolation. On a broader definition they may also include facts about relationships between individuals as well as the state of their environment. Different purportedly individualistic theories can differ in how restrictive or permissive their descriptive concepts and categories are.

The appeal of supervenience individualism should be clear: unless we employ an unreasonably restrictive criterion of what counts as an individual-level fact, it is hard to imagine a hypothetical duplicate of our world in which all individual-level facts remain

[9] If we were to define physical facts so broadly as to subsume *all* facts, then supervenience physicalism would be trivially true. But once we define physical facts more narrowly, for example to include only facts describable in the language of fundamental physics, supervenience physicalism is an interesting and substantial thesis.

unchanged but which nonetheless differs in some social facts. Indeed, it is hard to see how there could be differences at the social level that could not be traced to some differences, however subtle, at the individual level. Accordingly, supervenience individualism seems to be an important "no mystery" constraint on social facts (see also List and Pettit 2011). The negation of supervenience individualism would be a very radical – and we think implausible – form of "social-fact holism":

> **Social-fact holism:** The individual-level facts do not fully determine the social facts; i.e., there can be possible worlds that are identical with respect to all individual-level facts but different with respect to some social facts.

A thesis about particular objects

A second physicalist thesis concerns *particular objects*. Any scientific theory has some ontological commitments: it takes some kinds of objects as "ontologically primitive". Particle physics has electrons, neutrons, and protons in its ontology, for instance, while psychology has an ontology of mental states, beliefs, intentions, and so on. Furthermore, the objects in a theory's ontology need not be restricted to *entities* in the ordinary sense, but may also include *events* or *processes*. Mental events or processes are objects in the ontology of cognitive science, just as historical events or processes are objects in the ontology of history or political science. The physicalist thesis about particular objects is:

> **Token physicalism:** Every particular object in the psychological-level ontology is identical to ("the same as") some object in the physical-level ontology. (There is at most a difference in description.)

The idea is that every entity, event, or process in the ontology of psychology – say, a particular intentional attitude, perception, or thought process – coincides with some physical entity, event, or process, notwithstanding its conventional description in psychological terms. In principle, one could accept token physicalism with respect to some kinds of objects, say events, while rejecting it with respect to others, say entities or processes. The popular research programme of trying to find neural correlates for all sorts of mental phenomena reflects a token-physicalist view. Here, brain-imaging techniques are used to identify the brain state or neural process with which any particular mental state or process – for instance, a particular experience of pain or visual perception – coincides. The social-scientific counterpart of token physicalism is:

> **Token individualism:** Every particular object in our social ontology is identical to ("the same as") some object in the individual-level ontology. (There is at most a difference in description.)

For this to be true, every particular entity, event, or process that we conventionally describe as "social" must admit of a re-description in individual-level terms. Thus any entity such as a state, political party, or university must be identical to a specific collection or configuration of individuals, and any social event or process such as the 2008 US presidential election or the 2011 EU bailout of Greece must be identical to some event or process that can be fully described in individual-level terms. As with token physicalism, one could accept token individualism with respect to some kinds of objects, such as events, without accepting it with respect to others, such as entities or processes.

Whether token individualism with respect to any such objects is defensible depends on how narrowly or broadly these are construed. If we take a snapshot of a university or

nation at a given point in time, then perhaps it is not absurd (though still demanding) to identify it with some collection or configuration of individuals at that time. Similarly, if we construe a particular bailout as a specific sequence of individual actions, then the token-individualist stance may seem theoretically viable.

But once we view a social entity as something that can persist over time, or a social event or process as something that could have occurred under different circumstances, any such identification becomes a stretch. An indefinite number of individual-level constellations might give rise to what counts as the same social entity, event, or process.[10] An entity such as Facebook clearly occurs in our social ontology; yet, it seems practically impossible to pick out any entity described at the individual level with which it *exactly* coincides. If this observation is correct, it may support a form of "token holism":

[10] The more we view a given social entity as extending beyond a snapshot of a particular group of individuals at a time, or a social event or process as something that could have occurred via different individual actions, the less plausible the token-individualist stance becomes. Thus, higher-level particulars need not be identical to lower-level particulars. The University of London (a social particular and token of the type "university") might be multiply realizable at the individual level *once we admit its persistence through time and across different hypothetical circumstances*. Here we depart from Sawyer's acceptance of token identity, which he defines with respect to events (2002, 2003).

Token holism: Some particular objects in our social ontology are distinct from (and not re-describable as) any objects in the individual-level ontology.[11]

A thesis about properties

A third and even more demanding physicalist thesis concerns *properties*. Scientific theories refer not only to *particular objects*, such as particular entities, events, or processes, but also to *general properties* of those objects. Property ascriptions play important roles in explanations, as discussed below. Examples of physical properties are "having an electrical charge" or "having a certain momentum", while examples of psychological properties are "believing or desiring something", "experiencing fear or hope", "perceiving an object", and so on. Physicalism about properties is the following thesis:

Type physicalism: Every psychological property is identical to some physical property. (There is at most a difference in description.)

This is a demanding claim. General properties can usually be instantiated in myriad ways, and it may be impossible to spell out *in physical terms* what, for example, all the different brain states that may encode the belief that Kennedy was assassinated in 1963 have in common. The social-scientific counterpart of type physicalism is:

---

[11] This and subsequent holist theses have weaker and stronger interpretations, depending on what we take "not re-describable" to mean. Possible interpretations range from "not even in principle re-describable" to "not easily re-describable in practice".

Type individualism: Every social property is identical to some individual-level property. (There is at most a difference in description.)

A paradigmatic example of a social property may be "owning 20 dollars". To see how demanding it is to claim that this property is identical to some individual-level property, consider all the different ways in which someone may instantiate it. One may have certain slips of paper or coins in one's pocket or somewhere else. One may have a bank account, which, in turn, can be instantiated in many different ways (in the past, it was registered in a physical book, while today it is stored in a computer system). More generally, one may stand in a particular contractual relationship with others, which can again take innumerably many different forms. Even if every *token instance* of the property of owning 20 dollars could be translated into a complex individual-level configuration, such a translation is bound to fail when it comes to the *general property* itself (for detailed arguments, see Fodor 1974 and Sawyer 2002). This supports:

Type holism: Some social properties are distinct from (and not re-describable as) any individual-level properties.

A thesis about causal explanations

A final physicalist thesis whose social counterpart we wish to consider concerns causal explanations:

Causal-explanatory physicalism: Every causal relation (of the kind that a scientific explanation would describe) is identical to some physical causal relation. (There is at most a difference in description.)

The idea is that genuine causal relationships, which "rock-bottom explanations" should identify, exist only at the physical level. Apparent higher-level causal relations, on this picture, are at most "epiphenomenal". For example, when someone forms the intention to lift his or her arm and then acts on this intention, the cause of the resulting bodily movement is not the person's intention (a higher-level, mental state) but rather his or her brain state (a lower-level, physical state). The so-called "exclusion argument", to which we return in the next section, asserts that, in a causally closed physical world, all genuine causal relationships are physical ones. The corresponding social-scientific thesis is:

> **Causal-explanatory individualism:** Every causal relation (of the kind that a social-scientific explanation would describe) is identical to some individual-level causal relation. (There is at most a difference in description.)

An implication of this thesis is that any higher-level social patterns or regularities that we observe can be interpreted only as correlations and not as genuine causal relationships unless we have identified the underlying individual-level mechanisms. When we see evidence for an inverse relationship between the rate of unemployment and inflation in an economy, for example, we are only entitled to interpret this relationship, the so-called "Phillips curve", as a correlation. In the absence of micro-level foundations, we cannot legitimately claim that there is a causal connection between the two. This view is widely accepted, though seldom critically scrutinized, in political science and microeconomics. In the next section, we examine the case for causal-explanatory individualism in more detail. The negation of the thesis, which we come to defend, is:

Causal-explanatory holism: Some causal relations (of the kind that a social-scientific explanation would describe) are distinct from (and not re-describable as) any individual-level causal relations.

How are the different theses logically related to each other?

The need to distinguish between the different variants of each of individualism and holism is underscored by the fact that they are far from equivalent to one another, and that although each individualist thesis has a holistic negation, some individualist theses are consistent with other holistic ones. The logical relationships are as follows:

Proposition: (i) Supervenience individualism does *not* imply any of the other three individualist theses; (ii) supervenience individualism is logically independent from (especially, neither implies nor is implied by) token individualism; (iii) type individualism implies each of supervenience individualism, token individualism, and causal-explanatory individualism.

Corollary: (i) Social-fact holism is *not* implied by any of the other holistic theses; (ii) social-fact holism is logically independent from (especially, neither implies nor is implied by) token holism; (iii) type holism is implied by each of social-fact holism, token holism, and causal-explanatory holism.

In an online appendix, we sketch an analytic argument for all of these claims (in line with Stoljar 2009). We now focus on the most central claim for our reconciliation of individualism and holism (included in part (i) of the proposition and its corollary):

The compatibility thesis: Supervenience individualism is compatible with causal-explanatory holism.

Why supervenience individualism is compatible with causal-explanatory holism

As already noted, methodological individualists typically derive the conclusion that all social phenomena are in principle explicable in individualistic terms from the premise that the social world is the result of many individuals interacting in various ways. Thus they assume, contrary to our compatibility thesis, that supervenience individualism implies causal-explanatory individualism (see also Kincaid 1986). Since the argument for this claim is seldom carefully articulated (notable exceptions include Sawyer 2003 and Greve 2012),[12] it is worth reconstructing it and showing where it fails.

The exclusion argument

The argument is a version of what philosophers call the "exclusion argument" (e.g., Kim 1998, 2005). Consider a system that can be described at multiple levels:

(1) a lower level, e.g., the level of the brain in the mind-body case or the level of individuals in the social case; and

(2) a higher level, e.g., the level of the mind or the social level.

Let us further accept supervenience physicalism (in the mind-body case) or supervenience individualism (in the social case), i.e., all higher-level facts supervene on lower-level facts. Suppose we wish to explain a particular property of the system – call it E for "effect" – by identifying the property or combination of properties that causes it. We face the following methodological question:

[12] Sawyer discusses and responds to the argument; Greve criticizes his response. We find neither Sawyer's defence of social causation nor Greve's critique fully satisfactory. Here we offer an alternative response, based on the difference-making account of causation.

> **The key methodological question:** Should we identify the cause of an effect E at the lower level of description (e.g., a brain state in the mind-body case or a set of individual actions in the social case) or at the higher level (e.g., a mental state or an aggregate social pattern)?

In response to this question, two principles are frequently invoked:

> **The causal closure of the lower level (e.g., physical or individual):** Any effect E (regardless of whether E is described at the lower level or at the higher level) must ultimately have a lower-level cause; call it C.

> **The exclusion principle:** If a *lower-level* property C is the cause of E, no distinct *higher-level* property C* that supervenes on C can also be a cause of E. (So C* is at most an "epiphenomenon" of the "real" cause.)

From these two principles, it follows that the cause of E must be some lower-level property C, rather than some distinct higher-level property C* that supervenes on it.

Suppose, for example, we wish to explain why immigration queues at London Heathrow Airport have got longer – an effect – and we cite the government's tightening of immigration regulations – a higher-level property – as the cause. The causal-closure and exclusion principles would imply that this is the wrong causal attribution or just an epiphenomenon of the real cause. The real cause must be a complex set of individual actions by various officials leading to the immigration officers' extensive checks of each arriving passenger, which in turn lead to increased waiting times. On this picture, the higher-level property serves only as a "shorthand" for the real, micro-level cause on which it supervenes. In sum:

The argument: Given supervenience individualism, the causal-closure and exclusion principles imply (a version of) causal-explanatory individualism.

Since we accept supervenience individualism as a "no mystery" constraint on social facts, we can deny the conclusion of this argument only by denying either the causal-closure principle or the exclusion principle or both. Whether these withstand scrutiny depends on what we mean by a "cause". Drawing on recent work in philosophy (List and Menzies 2009; Menzies and List 2010), we now argue that if we define causation in the way that is most common in the social sciences, neither of the two principles is generally defensible.

Two conceptions of causation

Broadly speaking, there are at least two approaches to defining causation (e.g., Hall 2004): the "production- or mechanism-based" approach and the "difference-making" one (sometimes called "counterfactual" approach). Both can be seen as attempts to elucidate the traditional notion of "cause and effect", but they do so in very different ways. The first focuses on the *processes* or *mechanisms* "producing" certain effects, the second on the *regularities* in which certain events stand in actual and counterfactual conditions.

The production- or mechanism-based approach is best illustrated by the traditional idea that causation paradigmatically involves physical objects or bodies impacting on one another, transmitting forces, and thereby pushing one another around. (Think of a billiard ball colliding with another.) Thus causation is understood as a process or mechanism that produces certain outcomes. Developing this idea further, Russell (1913) argued that modern science, such as Newton's physics, does not "carve up" physical processes into discrete events (such as discrete "causes" and "effects") but describes continuous functional relations that capture "general causal laws": mechanisms by which systems

evolve over time, for instance when objects, such as the planets in the solar system, interact under various conditions.[13] Social-scientific approaches that emphasize the step-by-step individual-level mechanisms by which social phenomena are "produced", such as the micro-foundations programme in macro-economics, are best interpreted as following a production- or mechanism-based approach (Tilly 2001; Hedström and Ylikoski 2010).

The difference-making approach, by contrast, defines causation not in terms of processes or mechanisms, but in terms of the regularities in which certain events or event-types stand. This approach is particularly useful in many special (i.e., non-physical) sciences, especially when intentional decision-making or other higher-level phenomena are involved. In sciences ranging from medicine and ecology to political science and economics, we are often interested in how changes in some "independent" (or "causal") variables (e.g., through interventions) systematically relate to changes in certain "dependent" (or "effect") variables. On the difference-making approach, causal relationships are robust regularities between certain variables or properties (Lewis 1973; Brady 2008; Woodward 2009; see also Fearon 1991). So, a property C (within a system of interest) is the cause of another property E if and only if C systematically makes a difference to E. In the illustrative case in which C and E are binary properties, this can schematically be spelt out in terms of two conditionals:

> **Positive tracking:** If C occurred (other things being equal), E would occur. Formally, E is present in all closest possible worlds (relative to the actual world) in which C is present.

[13] For Russell, calling these laws "causal" is gratuitous; we could equally call them the "laws of physics" or the "laws of nature".

> **Negative tracking:** If C did not occur (other things being equal), E would not occur. Formally, E is absent in all closest possible worlds (relative to the actual world) in which C is absent.

The way in which we obtain evidence for causation in statistics, for instance in regression analysis, is very much in line with this way of thinking about causation: if causation is defined as difference making, then robust correlations – controlling for sufficiently many other factors – are indeed *evidence* for causation.

A response to the exclusion argument

The exclusion argument is most compelling when causation is understood in the production- or mechanism-based way (as intended in Kim 1998, 2005), at least to the extent that we are ultimately interested in the most fine-grained mechanisms underlying any given phenomenon. So it comes as no surprise that mechanism-based approaches to the social sciences, such as the micro-foundations programme in macroeconomics, are among the most strongly individualistic ones. When causation is understood as difference making, by contrast, it is an empirical question whether the most systematic regularities in which some effect E in a social system stands can be found at the lower level or at the higher level of description.

As just noted, scientists usually try to find evidence for causal relations by looking for robust correlations that persist even in the presence of careful experimental or statistical controls. Suppose we find a robust correlation between a higher-level "cause" property (e.g., the interest rate set by the central bank) and some "effect" property (e.g., inflation) in a system (here the economy). We can then ask whether the effect would continue to occur across variations in the lower-level realization of its putative higher-

level cause. If the effect continues to occur under at least some such variations (other things being equal), we call the higher-level causal relation "robust to changes in its microrealization", for short "microrealization-robust". While causal relations that lack such robustness are reducible to a lower level of description, robust ones not. In the latter case, the two "tracking conditionals" for difference-making causation are satisfied by the higher-level properties, *not* by the underlying lower-level properties (as shown in List and Menzies 2009; their term for "microrealization-robustness" is "realization-insensitivity").

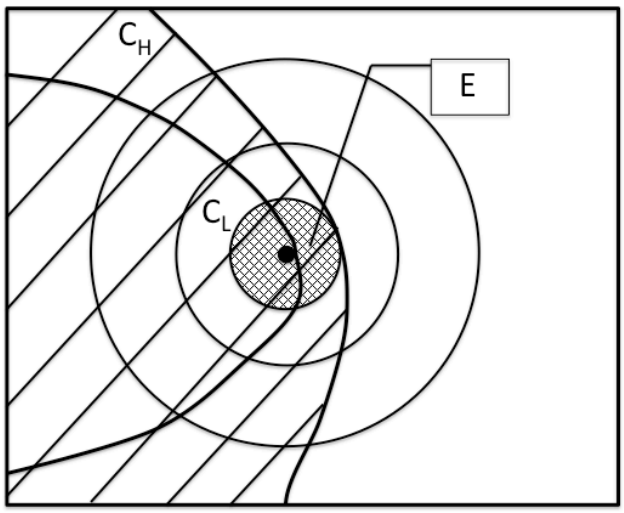

Figure 1: Microrealization-robust higher-level causation

Figure 1 (adapted from List and Menzies 2009) gives a schematic illustration. Here each point in the two-dimensional space represents one possible world (a "possible scenario"), with the central dot standing for the actual world (the "actual scenario"). The concentric circles around it include increasingly distant possible worlds, with fewer and fewer *ceteris paribus* conditions met. Suppose we want to explain what caused some effect E that actually occurred and would have occurred throughout the innermost circle of possible worlds (the shaded region). One candidate cause of E is the higher-level property C_H, which we find in all possible worlds in the large shaded region on the left. The two tracking conditionals introduced above are satisfied by C_H with respect to E.

Positive tracking is satisfied because in the closest possible worlds (relative to the actual world) in which C_H is present, E is present too; and negative tracking is satisfied because in the closest possible worlds in which C_H is absent (moving towards the right), E is absent too. Suppose now that the lower-level realization of the higher-level property C_H in the actual world is the lower-level property C_L. Since this is only one of many possible realizations of C_H, it is not present throughout the entire shaded region corresponding to C_H, but only in the smaller half-oval inside it. If one were to accept the logic of the exclusion argument, one would have to identify C_L as the cause of E and regard C_H as merely epiphenomenal. But only one of the two tracking conditionals – the positive one – is satisfied by C_L with respect to E: E is present in all closest possible worlds in which C_L is present. Negative tracking is violated, because it is not true that in the closest possible worlds in which C_L is absent, E is absent too: rather, E continues to occur even if the higher-level property C_H has a different lower-level realization. Accordingly, we have an instance of microrealization-robust causation.

As a stylized real-world example, consider the failed Copenhagen climate summit in 2010. Leaked audio recordings give a detailed picture of how the negotiations faltered in the final hours. An offended Wen Jiabao had withdrawn to his hotel room, directing his chief negotiator by mobile phone. Nicolas Sarkozy lost his temper. Angela Merkel was sidelined. Finally, Mammohan Singh, Lula da Silva, and others struck a minimal deal with Barack Obama (Rapp, Schwägerl, and Traufetter 2010). Fascinating as this individual-level account is, it is of limited use as a causal explanation. As many analysts have observed, the summit failed for structural reasons, such as the large number of parties, the lack of a common interest, the arguably unhelpful legal framework of the UNFCCC process, and a US president without a domestic mandate for any substantial

concessions (see, e.g., Victor 2011, ch. 7). The suggested higher-level causal relations are microrealization-robust, because the summit would have been likely to fail even if Wen Jiabao had not sent his chief negotiator, Nicolas Sarkozy had reined himself in, or a different group of politicians altogether had negotiated the same subject matter in broadly similar conditions.

The bottom line is that, once causation is understood as difference making, the causal-closure and exclusion principles are by no means conceptual truths about causation, but rather contingent principles that may apply to some causal systems but not to others. On which side of this divide a given system falls depends on whether it exhibits microrealization-robust higher-level regularities (List and Menzies 2009). We return to this point when we spell out criteria for identifying systems that require causal-explanatory holism. In sum, supervenience individualism does *not* imply causal-explanatory individualism, but is fully consistent with causal-explanatory holism.

Individualism and holism in political science

The debate between individualists and holists in political science often suffers from a lack of clarity, due to a conflation of the supervenience, token, type, and causal-explanatory dimensions. We now revisit some salient areas of that debate, beginning with the heated arguments for and against rational-choice approaches.

Rational choice theory and political economy

As Laver (1997, p. vii) recalls,

> "[t]here was a time when, if an academic colleague (most often a philosopher or sociologist) accused someone of being a reductionist then that was the end

> of the matter… I too was wounded by accusations of reductionism back in the 1980s but now I can honestly say that I really do see nothing wrong with it. Many more years since then of reading volumes of 'non-reductionist' writings on politics have not convinced me of their superior virtues."

But what exactly is the accusation levelled against Laver? Our taxonomy allows us to ask more specific questions: is Laver accused of subscribing to supervenience individualism, token individualism, type individualism, causal-explanatory individualism, or a combination of these positions? Laver does not make this precise, but the charge of "reductionism" gives a hint. First, one can set aside the supervenience thesis because, as we have seen, supervenience individualism does not entail the reducibility of all social-level to individual-level explanations and even Durkheimian type holists could accept it. Second, token individualism by itself would not imply any interesting form of explanatory reducibility either: if all particular objects in our social ontology – for example, particular events – could be re-described as individual-level objects, we would certainly be able to perform a reduction of tokens, but this would tell us little about social regularities in general. What is at stake between Laver and his critics is the reduction of types and causal explanations. The charge is *either* that Laver is a type individualist, who holds that all social properties are identical with individual-level properties; *or*, more plausibly (since type individualism is implausible), that he is a causal-explanatory individualist, who holds that good explanations should be stated in terms of individuals alone. This is a common position among rational choice theorists. But as we have argued, causal-explanatory individualism fails when the most robust regularities can be found at a higher level of description.

The individualist focus of rational choice theory goes back to some of its founders. Buchanan and Tullock (1962) describe a methodological and normative black-and-white choice between an "organic conception" of the state on the one hand, and taking the individual as "the primary philosophical entity" on the other (p. 11), warning that organic conceptions "postulate the emergence of a mystical general will that is derived independently of the decision making process" among individuals (p. 12). The unpalatable mysticism of social-fact holism is Buchanan and Tullock's argument for methodological individualism. But the black-and-white choice is too stark: as we have seen, the falsity of social-fact holism does not entail the falsity of other forms of holism.

The view that the economic analysis of politics requires purely individualistic explanations is not shared by everyone, even on the rational-choice side. Satz and Ferejohn (1994), for example, argue that there are different kinds of explanations, some of which operate at a micro-level, offering concrete psychological accounts of action, while others operate at a macro-level, focusing on aggregate behavioural patterns (see also Scharpf 1997). The upshot is that rational choice theory, broadly conceived, is not wedded to causal-explanatory individualism. An example of a broadly rational-choice-theoretic approach that embraces type and causal-explanatory holism is the emerging field of network analysis (e.g., Christakis and Fowler 2009), to which we return later.

Contexts and institutions

There are many areas of political science in which reference to higher-level aggregates, institutions, parties, cultures, groups, and social contexts is natural to the subject matter at hand, entailing an implicit commitment to causal-explanatory holism. For example, "[t]he central tenet of modern comparative politics is ... that *context* – structural, cultural,

institutional, and strategic; social, economic, and political; international, domestic, and local – matters" (Franzese 2009, p. 29). All the explanatory variables mentioned here refer to social properties. To see the ontological commitments of this approach, suppose for a moment that type individualism were true. This would imply that all social properties are identical to individual-level properties and thus that all those "context" properties could be reduced to individual-level properties. Clearly, that is not what most scholars of comparative politics believe. For many, it is precisely the irreducibility of structural, social-level properties that makes comparative analysis a worthwhile mode of investigation. Such a commitment to irreducible "context" variables requires type holism. If one also assumes that irreducible social-level properties feature in the best explanations (as Franzese suggests), then one arrives at causal-explanatory holism.

Similarly, the turn towards a "new institutionalism" has revived the analysis of political phenomena in terms of the "relative autonomy of institutions" rather than the "aggregate consequences of individual behaviour" (March and Olsen 1984, p. 734; cf. Peters 2012). Researchers using social properties as independent variables do not, however, need to assume the existence of mysterious "wholes" (and many new institutionalists are concerned with providing individual-level foundations; see, e.g., North 1993). As noted above, type and causal-explanatory holism are compatible with supervenience individualism and the more plausible, weak forms of token individualism.

States and other collective entities

In another area of political science, international-relations theory, the question of whether one should take only individuals as the ultimate unit of analysis is particularly pressing, because a common unit of explanation is the state. A type individualist with regard to

states believes that all properties of a state are nothing but properties of the individuals constituting it, and a causal-explanatory individualist believes that all explanations in international-relations theory can be given in terms of individuals alone, without treating states as actors in their own right. If type and causal-explanatory individualism were true, the sub-discipline of international-relations theory could be reduced to theories of individual behaviour. Resisting this view, and drawing, like us, on non-reductive physicalism in the philosophy of mind, Wendt (2004) argues that a commitment to the individual as the basic unit of our ontology does not imply that a reduction from state-level to individual-level explanations is possible. This is in line with our proposition that supervenience individualism does not imply any of the other forms of individualism.

Questions about the ontological status, and explanatory significance, of higher-level units can also be asked about other collective entities in politics, such as legislatures, committees, parties, interest groups, and non-governmental organizations. Hay (2006, pp. 88/89), for example, writes:

> "In political analysis and the philosophy of the social sciences more broadly there is no more hardy perennial than the question of the relationship between individuals and social collectivities or groups... Can collective actors (states, political parties, social movements, classes, and so forth) realistically, or indeed just usefully, be said to exist? If so, do they exhibit organic qualities, such that their character or nature is not simply reducible to the aggregation of the constituent units (generally individual actors) from which they are forged? Are such entities … appropriate subjects of political analysis and, if so, what if any behavioural characteristics can be attributed to them?"

Recently, there has been a growing interest in the thesis that the most parsimonious explanations of the behaviour of such collectivities often involve modelling them as purposive, rational agents in their own right, with preferences and judgments that need not be a simple function of the underlying individual preferences and judgments, although they supervene on them (e.g., Pettit 2001, ch. 5, 2003; Tollefsen 2002; List and Pettit 2002, 2011). A central idea is that the higher-level regularities in the behaviour of collective entities may sometimes warrant taking what Dennett (1987) calls an "intentional stance" towards them. Even micro-economists take this stance towards some collectives, for example when they model firms as unified rational actors in the theory of the firm. Many, of course, would argue that the ascription of agency to firms is nothing but an instrumentally useful modelling simplification, and the notion of group agency remains controversial (for a critical discussion, see, e.g., Kornhauser 2008).

A new consensus?

Hay (2006, p. 90) suggests that in recent years a "commonsense ground" has emerged within political science. It supposedly consists of the claims that (i) there are irreducible social wholes, (ii) these wholes have properties of their own, which cannot sensibly be seen as properties of their constituent parts, but (iii) the wholes are ultimately constituted only by their parts, so that there is no mysterious additional ontological ingredient. In our terms, Hay accepts supervenience individualism, type holism, and causal-explanatory holism, and does not take a stand on the token dimension. The conjunction of (i), (ii), and (iii) also echoes Sawyer's "non-reductive individualism" (2002, 2003).

We are not sure whether this "commonsense ground" is as entrenched in the discipline as Hay suggests; the short survey in this section raises some doubts. Perhaps a

general consensus on the matter is not even desirable, as different phenomena may require different methodological approaches. We now suggest criteria to determine which systems require "holist" methodologies and which require "individualist" ones.

Systems requiring holistic versus individualistic methodologies

We have shown that causal-explanatory holism is plausible when the systems or phenomena in question display robust causal regularities – in the difference-making sense – at the higher, aggregate level, but not at the lower, individual level. Applying our analysis, we characterize such systems or phenomena in terms of three formal conditions and give some political-science illustrations.

Three conditions for causal-explanatory holism

A social system requires explanatory holism if and only if three jointly necessary and sufficient conditions are met:

> **Multiple levels of description:** The system admits lower and higher levels of description, associated with different level-specific properties (e.g., individual-level properties versus aggregate-level properties).
>
> **Multiple realizability of higher-level properties:** The system's higher-level properties are determined by its lower-level properties, but can be realized by numerous different configurations of them and hence cannot feasibly be re-described in terms of lower-level properties.

> **Microrealization-robust causal relations:** The causal relations in which some of the system's higher-level properties stand are robust to changes in their lower-level realization.

The first condition is almost always met by social systems or phenomena. The lower level of description typically refers to individuals and their properties, while the higher level refers to the properties of social aggregates.

The second condition demands that although higher-level facts supervene on lower-level facts, many configurations of lower-level properties can instantiate the same higher-level properties. This multiple realizability implies that higher-level properties are equivalent at most to complicated enumerations of lower-level properties ("wild disjunctions", as discussed in Sawyer 2002, 2003). Evidently, many social-level properties (e.g., an unemployment rate of 8%, a single-peaked profile of preferences, a conservative majority in parliament etc.) can be realized by many different configurations of individual-level properties. Think of all the different possible distributions of jobs and job-seekers that would correspond to an unemployment rate of 8%, the different possible preference profiles that are single-peaked, and the different possible conservative winning coalitions of legislators.

With regard to the third condition, recall that when a higher-level property is a difference-making cause of another property (satisfying positive and negative tracking), this causal relation is "microrealization-robust" if the effect property would continue to occur under at least some variations in the lower-level realization of its putative cause. As we have seen, while causal relations that lack this robustness can be reduced to the lower level of description, this is not the case for microrealization-robust causal relations (List

and Menzies 2009). Hence higher-level descriptions are indispensable in causal explanations of systems involving microrealization-robust causal relations.

The study of ethnic conflicts and civil war

To illustrate these conditions, consider first one of the most cited articles in the *American Political Science Review*, Fearon and Laitin's "Ethnicity, Insurgency, and Civil War" (2003). Their central claim is that

> "[t]he conditions that favor insurgency – in particular, state weakness marked by poverty, a large population, and instability – are better predictors of which countries are at risk for civil war than are indicators of ethnic and religious diversity" (p. 88).

Note first that the phenomena of interest clearly admit *multiple levels of description*. We can describe them either at the aggregate level of countries or groups, or at the level of individuals. Although Fearon and Laitin occasionally offer rough sketches of what a lower-level mechanism might look like (e.g., p. 79), they are interested in identifying causal relations between higher-level properties. Second, *multiple realizability of higher-level properties* is also satisfied. While facts about state weakness or insurgency supervene on certain individual-level facts (the totality of all facts about individuals determines the aggregate facts about states, populations, etc.), these aggregate facts can nevertheless be realized in so many different ways that it may be hard to enumerate all their possible realizing conditions. For example, an aggregate property such as "state weakness" can be instantiated by myriads of different individual constellations, whose enumeration is neither practically feasible nor useful. Third, Fearon and Laitin plausibly

discover *microrealization-robust causal relations*, since the identified higher-level regularities are most likely to be robust to changes in individual-level realization. Just as the Copenhagen climate summit would probably still have failed with different political actors in place, state weakness will cause insurgency independently of how exactly it is realized. If Fearon and Laitin are right, the causal relation between state weakness and insurgency should hold independently of the details of the realizing individuals, since the causal relation is supposed to hold not only for past but also for future cases, for which individual realizations are unknowable.

Social-network theory

As a second example, consider the newly emerging field of network analysis. Christakis and Fowler (2009, p. 32) suggest that

> "[i]f we want to understand how society works, we need to fill in the missing links between individuals. We need to understand how interconnections between and interactions between people give rise to wholly new aspects of human experience that are not present in the individuals themselves."

Assume, for instance, that a property of the network structure (e.g., average path length or the property of being "scale-free") determines how quickly political information is disseminated. There are *multiple levels of description*, namely individuals (network nodes) on the one hand, and aggregate network properties (the structure of links) on the other. Identical network properties can be realized by many different constellations of individuals, which implies that some *higher-level properties are multiply realizable*. Most importantly, higher-level causal relations (based, for instance, on a correlation

between average path length and the spread of information) are plausibly *microrealization-robust.* The point of the network-theoretic research programme is to identify structural properties of networks that feature in certain causal relations, even though the specific networks instantiating them can differ dramatically. For instance, two networks can have the same aggregate property (such as being "scale-free") while having very different local topologies and nodes. It is the property of being scale-free that is robustly correlated with certain effects, and this correlation does not depend on the precise local topology or the properties of individual nodes.

States in international relations

As already noted, international-relations scholars face the question of what the appropriate unit of analysis in explanations of international politics should be: individuals, governments, states, or even larger units? Here again, our three conditions for causal-explanatory holism are typically met. First, there are *multiple levels of description*, because one can describe phenomena in international relations at the state level, individual level, or various sub- or supra-national levels. Second, *multiple realizability of higher-level properties* holds because the same behaviour of a state can be realized in many different ways at the individual level. Third, if international relations theory ever uncovers genuine causal relations that go beyond single cases, these must almost by definition be *microrealization-robust*. Consider for instance the "democratic peace hypothesis", according to which "democracies do not fight each other because norms of compromise and cooperation prevent their conflicts of interest from escalating into violent clashes" (e.g., Maoz and Russett 1993, p. 624). The higher-level property of "being a democracy" supervenes on the properties of individuals; i.e., it must ultimately

be instantiated by a complex mix of individual attitudes, beliefs, behaviours, and so on. However, listing all possible individual-level constellations that can instantiate "being a democracy" is practically impossible. Furthermore, the point of the democratic peace hypothesis is that there are structural features of democracies that lead them to avoid wars with one another, independently of the individual political details of each case (e.g., the personalities of the relevant politicians, the specifics of the public debate on the issue, the voting pattern in the legislature, and even the details of the democratic constitution itself). For this reason, if there is a causal relation between "being a democracy" and "avoiding wars with other democracies", it must be microrealization-robust.

Concluding remarks

Much of the stalemate in the debate about individualism and holism in political science is arguably due to (i) a lack of precision in the definition of individualism and holism, (ii) the failure to distinguish sufficiently clearly between different variants of each view, (iii) a conflation of ontological and explanatory claims, and (iv) some confusion over whether supervenience implies explanatory reducibility, and whether causal-explanatory holism commits us to any more radical form of social-fact holism. By importing and developing ideas from the philosophy of mind and the philosophy of social science, we have sought to address each of these issues.

Most political scientists should be able to agree with supervenience individualism and reject the most radical form of social-fact holism. The choice between any of the other forms of individualism and holism – the type, token, and causal-explanatory ones – depends on what the object of study is. Token individualism is a demanding view, as evident in the difficulty of identifying the precise individual-level objects that entities

such as Facebook or the Supreme Court correspond to. However, even if we are inclined towards token individualism, the problem of multiple realizability may rule out type individualism and support type holism when it comes to devising descriptive categories for capturing higher-level social and political phenomena. Causal-explanatory holism, finally, is mandated when (i) we want to identify difference-making causes as opposed to maximally fine-grained mechanisms and (ii) higher-level regularities are microrealization-robust. Although we may expect microrealization-robust regularities to be common in many social phenomena, it is an empirical question where they occur, and empirical research may also uncover some regularities that depend on the details of their individual-level realization. For the latter, and for any research on maximally fine-grained causal mechanisms, our arguments support causal-explanatory individualism.

It is worth noting that even the proponents of causal-explanatory individualism are implicitly committed to at least one aspect of our case for non-reductive explanations. The human individual itself is a composite system consisting of trillions of cells, and so the assumption that the properties of individuals in their entirety, rather than those of their underlying cells, are significant in most social-scientific explanations must stem from the assumption that the relevant regularities are robust to changes in their microrealizations at the cellular level. What we are pointing out is that the case for non-reductive explanations need not stop at the level of the individual.[14]

We have focused on ontological and explanatory debates and have not discussed similar debates about individualism and holism concerning evaluative and normative matters. In those debates, the central questions are what or who the objects of (non-instrumental) value or moral concern should be – individuals or larger social entities, for

[14] We are grateful to an anonymous referee for prompting us to emphasize this point.

example – and who the bearers of rights and responsibilities should be. Although these questions are beyond the scope of this paper, the ontological and explanatory debates we have addressed can *inform* our views on them. For example, whether we recognize certain collectives – say corporations, political parties, or states – as agents from an ontological or explanatory perspective will be *relevant* to whether, normatively speaking, we consider them appropriate bearers of responsibility (e.g., French 1984; List and Pettit 2011). Similarly, for something to count as an object of value or moral concern, it must presumably feature in our all-things-considered defensible ontology. The framework we have developed should therefore be of interest to normative political theorists as much as to political scientists more generally.

References

Arrow, K. J. 1994. 'Methodological Individualism and Social Knowledge.' *American Economic Review* 84 (2): 1-9.

Brady, H. E. 2008. 'Causation and Explanation in Social Science.' In *The Oxford Handbook of Political Methodology*, eds. D. Collier, H. E. Brady, and J. M. Box-Steffensmeier. Oxford: Oxford University Press.

Bratman, M. E. 1999. *Faces of intention: selected essays on intention and agency*. Cambridge: Cambridge University Press.

Buchanan, J. M., and G. Tullock. 1967. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor: University of Michigan Press.

Chalmers, D. 1996. *The Conscious Mind*. Oxford: Oxford University Press.

Christakis, N. A., and J. H. Fowler. 2009. *Connected: the surprising power of our social networks and how they shape our lives*. New York: Little Brown.

Dawkins, R. 1986. *The blind watchmaker: why the evidence of evolution reveals a universe without design*. London: Penguin.

Dennett, D. 1987. *The Intentional Stance*. Cambridge, MA: MIT Press.

Dryzek, J. S. 2010. 'Rhetoric in Democracy: A Systemic Appreciation.' *Political Theory* 38 (3): 319-339.

Elster, J. 1982. 'The case for methodological individualism.' *Theory and Society* 11 (4): 453-482.

Elster, J. 1985. *Making Sense of Marx*. Cambridge: Cambridge University Press.

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

Epstein, J. M. 2006. *Generative Social Science*. Princeton: Princeton University Press.

Fearon, J. D. 1991. 'Counterfactuals and Hypothesis Testing in Political Science.' *World Politics* 43 (2): 169-195.

Fearon, J. D., and D. Laitin. 2003. 'Ethnicity, Insurgency, and Civil War.' *American Political Science Review* 97 (1): 75-90.

Fodor, J. 1974. 'Special sciences (or: the disunity of science as a working hypothesis).' *Synthese* 28 (2): 97-115.

Franzese Jr, R. J. 2009. 'Multicausality, Context-Conditionality, and Endogeneity.' In *The Oxford Handbook of Comparative Politics*, eds. S. C. Stokes and C. Boix. Oxford: Oxford University Press, pp. 27-72.

French, P. 1984. *Collective and Corporate Responsibility*. New York: Columbia University Press.

Gilbert, M. 2006. *A theory of political obligation: membership, commitment, and the bonds of society*. Oxford and New York: Oxford University Press.

Godfrey-Smith, P. 2009. *Darwinian populations and natural selection*. Oxford and New York: Oxford University Press.

Greve, J. 2012. 'Emergence in Sociology.' *Philosophy of the Social Sciences* 42 (2): 188-223.

Hall, N. 2004. 'Two concepts of causation.' In *Causation and Counterfactuals*, eds. J. Collins, N. Hall, and L. A. Paul. Cambridge, MA: MIT Press, pp. 225-276.

Hay, C. 2006. 'Political ontology.' In *The Oxford Handbook of Contextual Political Analysis*, eds. R. E. Goodin and C. Tilly. Oxford: Oxford University Press, pp. 78-96.

Hayek, F. A. v. 1942. 'Scientism and the Study of Society.' *Economica* 9 (35): 267-291.

Heath, J. 2011. 'Methodological Individualism.' In *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta. Available at: http://plato.stanford.edu/archives/spr2011/entries/methodological-individualism/

Hedström, P., and P. Ylikoski. 2010. 'Causal mechanisms in the social sciences.' *Annual Review of Sociology* 36: 49-67.

Hodgson, G. M. 2007. 'Meanings of methodological individualism.' *Journal of Economic Methodology* 14 (2): 211-226.

Jervis, R. 1997. *System Effects: Complexity in Political and Social Life*. Princeton: Princeton University Press.

Kim, J. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.

Kim, J. 2005. *Physicalism or Something Near Enough*. Princeton: Princeton University Press.

Kincaid, H. 1986. 'Reduction, explanation, and individualism.' *Philosophy of Science* 53 (4): 492-513.

Kitcher, P. 1984. '1953 and all that. A Tale of Two Sciences.' *The Philosophical Review* 93 (3): 335-373.

Kornhauser, L. 2008. 'Aggregate rationality in adjudication and legislation.' *Politics, Philosophy and Economics* 7 (1): 5-27.

Laver, M. 1997. *Private desires, political action: an invitation to the politics of rational choice*. Revised edition. London: Sage Publications.

Lewis, D. 1973. *Counterfactuals*. Oxford: Blackwell.

List, C., and P. Menzies. 2009. 'Non-Reductive Physicalism and the Limits of the Exclusion Principle.' *Journal of Philosophy* CVI (9): 475-502.

List, C., and P. Pettit. 2002. 'Aggregating Sets of Judgments: An Impossibility Result.' *Economics and Philosophy* 18 (1): 89-110.

List, C., and P. Pettit. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press.

Little, D. 1991. *Varieties of Social Explanation*. Boulder: Westview Press.

Lukes, S. 1968. 'Methodological Individualism Reconsidered.' *The British Journal of Sociology* 19 (2): 119-129.

Macdonald, G., and P. Pettit. 1981. *Semantics and Social Science*. London: Routledge.

Maoz, Z., and B. Russett. 1993. 'Normative and Structural Causes of Democratic Peace, 1946-1986.' *American Political Science Review* 87 (3): 624-638.

March, J., and J. Olsen. 1984. 'The new institutionalism: organizational factors in political life.' *American Political Science Review* 78 (3): 734-749.

Mansbridge, J., J. Bohman, S. Chambers, T. Christiano, A. Fung, J. Parkinson, D. F. Thompson, and M. E. Warren. 2012. 'A systemic approach to deliberative democracy.' In *Deliberative Systems*, eds. J. Parkinson and J. Mansbridge. Cambridge: Cambridge University Press, pp. 1-26.

Menzies, P., and C. List. 2010. 'The Causal Autonomy of the Special Sciences.' In *Emergence in Mind*, eds. C. Mcdonald and G. Mcdonald. Oxford: Oxford University Press, pp. 108-128.

Mill, J. S. 1974 [1843]. *A System of Logic: Ratiocinative and Inductive*. Toronto: University of Toronto Press.

North, D. 1993. 'Economic Performance through Time.' Lecture to the memory of Alfred Nobel, December 9, 1993. Available at: http://www.nobelprize.org/nobel_prizes/economics/laureates/1993/north-lecture.html

Ostrom, E. 2009. 'A General Framework for Analyzing Sustainability of Social-Ecological Systems.' *Science* 325 (5939): 419-422.

Parsons, T. 1937. *The Structure of Social Action*. New York and London: Macmillan.

Peters, B. G. 2012. *Institutional Theory in Political Science*. 3rd ed. New York and London: Continuum.

Pettit, P. 1993. *The Common Mind*. Oxford: Oxford University Press.

Pettit, P. 2001. *A Theory of Freedom*. New York: Oxford University Press.

Pettit, P. 2003. 'Groups with Minds of their Own.' In *Socializing Metaphysics*, ed. F. Schmitt. New York: Rowan and Littlefield, pp. 167-193.

Popper, K. R. 1944. 'The Poverty Of Historicism I&II.' *Economica* 11: 86-103, 119-37.

Popper, K. R. 1945a. *The Open Society and Its Enemies*. London: Routledge.

Popper, K. R. 1945b. 'The Poverty Of Historicism III.' *Economica* 12: 69-89.

Putnam, H. 1967. 'The Nature of Mental States.' In *Art, Mind, and Religion*, eds. W. H. Capitan and D. D. Merrill. Pittsburgh: University of Pittsburgh Press, pp. 37-48.

Rapp, T., C. Schwägerl, and G. Traufetter. 2010. 'How China and India Sabotaged the UN Climate Summit.' *Spiegel online*, 5 May 2010. Available at: http://www.spiegel.de/international/world/the-copenhagen-protocol-how-china-and-india-sabotaged-the-un-climate-summit-a-692861-3.html [acc. 26 Oct. 2012]

Rosenberg, A. 1997. 'Reductionism Redux: Computing the Embryo.' *Biology and Philosophy* 12: 445-470.

Russell, B. 1913. 'On the Notion of Cause.' *Proceedings of the Aristotelian Society* 13: 1-26.

Satz, D., and Ferejohn, J. 1994. 'Rational Choice and Social Theory.' *Journal of Philosophy* 91 (2): 71-87.

Sawyer, R. K. 2002. 'Nonreductive Individualism. Part I – Supervenience and Wild Disjunction.' *Philosophy of the Social Sciences* 32 (4): 537-559.

Sawyer, R. K. 2003. 'Nonreductive Individualism. Part II – Social Causation.' *Philosophy of the Social Sciences* 33 (2): 203-224.

Sawyer, R. K. 2005. *Social Emergence: Societies as Complex Systems*. Cambridge: Cambridge University Press.

Scharpf, F. W. 1997. *Games Real Actors Play: Actor-Centered Institutionalism in Policy Research*. Boulder: Westview Press.

Schumpeter, J. 1908. *Das Wesen und der Hauptinhalt der theoretischen Nationaloekonomie*. Leipzig: Duncker & Humblot.

Searle, J. R. 1995. *The Construction of Social Reality*. London: Penguin.

Searle, J. R. 2010. *Making the Social World: The Structure of Human Civilization*. Oxford and New York: Oxford University Press.

Smuts, J. C. 1926. *Holism and Evolution*. London: Macmillan.

Stoljar, D. 2009. 'Physicalism.' In *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta. Available at: http://plato.stanford.edu/archives/fall2009/entries/physicalism

Stoljar, D. 2010. *Physicalism*. Abingdon and New York: Routledge.

Tilly, C. 2001. 'Mechanisms in political processes.' *Annual Review of Political Science* 4 (1): 21-41.

Tollefsen, D. P. 2002. 'Collective Intentionality and the Social Sciences.' *Philosophy of the Social Sciences* 32 (1): 25-50.

Tuomela, R. 2007. *The Philosophy of Sociality: The Shared Point of View*. New York: Oxford University Press.

Tuomela, R. 2010. 'Holistic Social Causation and Explanation.' In *Explanations, Prediction, and Confirmation*, eds. D. Dieks et al. Heidelberg: Springer.

Udehn, L. 2001. *Methodological individualism: Background, history and meaning*. London and New York: Routledge.

Udehn, L. 2002. 'The Changing Face of Methodological Individualism.' *Annual Review of Sociology* 28: 479-507.

Vermeule, A. 2012. *The System of the Constitution*. Oxford: Oxford University Press.

Victor, D. G. 2011. *Global warming gridlock: creating more effective strategies for protecting the planet*. Cambridge and New York: Cambridge University Press.

Watkins, J. W. 1952. 'The principle of methodological individualism.' *The British Journal for the Philosophy of Science* 3 (10): 186-189.

Watkins, J. W. 1955. 'Methodological Individualism: A Reply.' *Philosophy of Science* 22 (1): 58-62.

Watkins, J. 1957. 'Historical explanation in the social sciences.' *The British Journal for the Philosophy of Science* 8 (30): 104-117.

Weber, M. 1978. *Economy and society*. Transl. G. Roth and C. Wittich. Vol. 1. Berkeley: University of California Press.

Wedeen, L. 2002. 'Conceptualizing Culture: Possibilities for Political Science.' *American Political Science Review* 96 (4): 713-728.

Wendt, A. 1999. *Social Theory of International Politics*. Cambridge: Cambridge University Press.

Wendt, A. 2004. 'The State as Person in International Theory.' *Review of International Studies* 30 (2): 289-316.

Woodward, J. 2009. 'Agency and Interventionist Theories.' In *The Oxford Handbook of Causation*, eds. H. Beebee, C. Hitchcock, and P. Menzies. Oxford: Oxford University Press, pp. 234-262.