Review of International Studies (2014), 40, 367–389
doi:10.1017/S0260210513000156
© 2013 British International Studies Association
First published online 17 Jun 2013

Mechanisms and microfoundations in International Relations theory

WILLIAM MCGINLEY

Abstract. This article looks at prospects for a mechanism-based research strategy in the study of International Relations. Over the past three decades, the notions of mechanism and microfoundation have taken a central place in discussions of explanation and 'micro-macro' problems in social science. The upshot of much of this discussion has been a call for mechanism-based explanations – explanations of macro-level phenomena in terms of micro-level mechanisms. Some work of this kind can already be found in IR theory, including in systemic research. However, a number of IR theorists, including Kenneth Waltz and Alexander Wendt, have argued that micro-oriented strategies like this will not work, pointing to incongruities between system- and unit-level phenomena. This article argues that these pose less hindrance to a fully-developed model of mechanism-based explanation, and that the field has much to gain from further exploration of this strategy. In particular, mechanism-based explanations could help bring structure back to the centre of discussion in IR theory, and might even give us a way out of the field's own micro-macro problems.

William McGinley is a PhD candidate in the School of Communication, International Studies, and Languages at the University of South Australia.

Introduction

When a theory explains something with reference to a large-scale cause, scientists will often look for what, at the micro-level, is doing the 'real' causal work. The same tendency is found in International Relations (IR) theory, even in system-level research. In work on the influence of systemic polarity and polarisation on war, for example, system-level statistical correlations are typically augmented with accounts of decision-making processes, which help explain the correlations themselves.[1] In constructivist research on the influence of institutional structure on agents, agent-level factors are frequently brought in to help explain how structure emerges and actualises its influence.[2] System-level research is not especially popular, but even those that do do it are rarely satisfied with macro-explanations alone.

[1] See, among many others, Karl W. Deutsch and J. David Singer, 'Multipolar Power Systems and International Stability', *World Politics*, 16:3 (1964), pp. 390–406; Bruce Bueno de Mesquita, 'Systemic Polarization and the Occurrence and Duration of War', *The Journal of Conflict Resolution*, 22:2 (1978), pp. 241–67; and Kenneth Waltz, *Theory of International Politics* (Long Grove, IL: Waveland Press, 1979), ch. 8. I elaborate on these and other examples below.

[2] See, for example, George M. Thomas, John W. Meyer, Francisco O. Ramirez, and John Boli (eds), *Institutional Structure: Constituting State, Society, and the Individual* (Newbury Park, CA: Sage, 1987); John Gerard Ruggie, *Constructing the World Polity: Essays on International Institutionalization* (London: Routledge, 1998); and Alexander Wendt, *Social Theory of International Politics* (Cambridge: Cambridge University Press, 1999).

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

What these and other theorists seem to be doing is looking for *mechanisms*, or *microfoundations*, in order to make better sense of causal work done by structure at the system or unit level. Over the past three decades, the notions of mechanism and microfoundation have taken a central place in discussions of explanation and 'micro-macro' problems in social science. Debate over ideals like 'holism' and 'individualism' has given way to the more concrete issue of whether structural phenomena can be given microfoundations. Those on the microfoundational side call for mechanism-based explanations – explanations of macro-level phenomena in terms of micro-level mechanisms.[3] Aided by the spread of the more general mechanistic worldview in philosophy and biology, mechanism-based explanation has become something of an institution in social science.[4] IR theory itself has been an active participant in this turn to mechanisms. The term mechanism has been used in the field for a long time, but over the past two or so decades its occurrence has increased noticeably, particularly in discussions of causation and explanation.[5] Scientific realists like David Dessler and Alexander Wendt have helped add depth to the notion, with their ontological accounts of causation.[6] Talk of microfoundations is also common, and reductionist theories of IR are increasingly understood in terms of this idea.[7]

This article looks at prospects for further exploration and development of the microfoundational strategy in IR theory. Because it involves linking phenomena at different levels of organisation, I argue, mechanism-based explanation might help resolve IR theory's own micro-macro problems – the level-of-analysis problem and the agent-structure problem. However, one finds few fully-fledged mechanism-based explanations in IR theory, and relatively little discussion of the methods involved,

[3] See, for example, Randall Collins, 'On the Microfoundations of Macrosociology', *The American Journal of Sociology*, 86:5 (1981), pp. 984–1014; Jon Elster, *Nuts and Bolts for the Social Sciences* (Cambridge: Cambridge University Press, 1989); Daniel Little, *Varieties of Social Explanation: An Introduction to the Philosophy of Social Science* (Boulder, CO: Westview Press, 1991); and Peter Hedström and Richard Swedberg (eds), *Social Mechanisms: An Analytical Approach to Social Theory* (New York: Cambridge University Press, 1998). See also James S. Coleman, *Foundations of Social Theory* (Cambridge, MA: Harvard University Press, 1990).

[4] On the mechanistic worldview, see John Gerring, 'The Mechanismic Worldview: Thinking Inside the Box', *British Journal of Political Science*, 38:1, (2007), pp. 161–79; and Peter Hedström and Petri Ylikoski, 'Causal Mechanisms in the Social Sciences', *Annual Review of Sociology*, 36:1, (2010), pp. 49–67, among others.

[5] See, for example, T. Clifton Morgan and Sally Howard Campbell, 'Domestic Structure, Decisional Constraints, and War: So Why Kant Democracies Fight?', *The Journal of Conflict Resolution*, 35:2 (1991), pp. 187–211; Lars-Erik Cederman, 'Modeling the Democratic Peace as a Kantian Selection Process', *The Journal of Conflict Resolution*, 45:4 (2001), pp. 470–502; Patrick Thaddeus Jackson and Daniel H. Nexon, 'Whence Causal Mechanisms? A Comment on Legro', *Dialogue IO*, 1:1 (2002), pp. 81–102; Stacie E. Goddard and Daniel H. Nexon, 'Paradigm Lost? Reassessing Theory of International Politics', *European Journal of International Relations*, 11:1 (2005), pp. 9–61; John A. Vasquez, *The War Puzzle Revisited* (Cambridge: Cambridge University Press, 2009); and Rudra Sil and Peter Katzenstein, 'Analytic Eclecticism in the Study of World Politics: Reconfiguring Problems and Mechanisms across Research Traditions', *Perspectives on Politics*, 8:2 (2010), pp. 411–31.

[6] See David Dessler, 'Beyond Correlations: Toward a Causal Theory of War', *International Studies Quarterly*, 35:3 (1991), pp. 337–55; and Wendt, *Social Theory*.

[7] Some that make reference to the idea include Andrew Moravcsik, 'Taking Preferences Seriously: A Liberal Theory of International Politics', *International Organization*, 51:4 (1997), pp. 513–53; Jack S. Levy, 'Prospect Theory, Rational Choice, and International Relations', *International Studies Quarterly*, 41:1 (1997), pp. 87–112; Jeffrey T. Checkel, 'The Constructivist Turn in International Relations Theory', *World Politics*, 50:2 (1998), pp. 324–48; Wendt, *Social Theory*; David A. Lake and Robert Powell (eds), *Strategic Choice and International Relations*, (Princeton: Princeton University Press, 1999); and David Pak Yue Leon, 'Reductionism, Emergence, and Explanation in International Relations Theory', in Jonathan Joseph and Colin Wight (eds), *Scientific Realism and International Relations* (New York and Basingstoke: Palgrave Macmillan, 2010), pp. 31–50.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

despite there being a fairly common inclination toward explanation of this kind. One also finds few explicit defences of the approach at the meta-theoretical level, despite it often being observed that micro-oriented research dominates the field.[8] Instead, one finds a substantial amount of criticism. In probably the best-known discussions of the issue, Kenneth Waltz criticised micro-oriented theory for its inability to account for the influence of the systemic context. In his most incisive discussion, he argued that there is a divergence in variation between unit-level causes and international outcomes, which renders the former explanatorily impotent.[9] These criticisms have been built on in recent times by a number of scientific realists. The charge has been headed by Alexander Wendt, who attacks microfoundationalism on both its explanatory and ontological fronts.[10] On the former, Wendt argues that structural factors are 'multiply realisable' at lower levels of analysis, making reductive explanation unfeasible. On the ontological front, he argues that structures are 'emergent' from their agent-level foundations, making any sort of reductive, microfoundational ontology incomplete.

The implication seems to be that mechanistic explanations of systemic phenomena will not be forthcoming, and we need to turn away from this downward-looking strategy, at least as a way of dealing with micro-macro linkages. I will argue, however, that the obstacles mentioned pose much less of a hindrance to a properly developed strategy of mechanism-based explanation – something the field lacks. I offer a systematic account of the strategy, building on work in biology and the philosophy of science. The aim of a mechanism-based explanation, on this account, is to account for the causal capacities of a macro-level system with reference to its components and their properties. The theorist begins with a functional description of the behaviour of a system, then searches for the micro-level mechanisms that realise this behaviour. Relative to this account, context ceases to be an obstacle. Rather, it is the very problem which mechanism-based explanations are designed to explain. Multiple realisability and emergence are more substantial obstacles, but not insurmountable. The former, I argue, can be handled through a strategy of localisation, or what Most and Starr refer to as 'nice laws'. The latter, I argue, can only really be established by the failure of mechanism-based explanations *in practice*, and not *a priori*, as some critical realists seem to suggest. The upshot of all this is that microfoundationalism does seem to be on the right track, and may very well give the field a way out of its micro-macro problems. But this comes with a twist. By linking structure with underlying mechanisms, mechanism-based explanations remove the mystery surrounding it, thus enhancing its credibility as an explanatory tool. Far from marginalising structural IR theory, the microfoundational strategy could give it much needed credibility.

The article is divided into three parts. In part one I look at some instances of microfoundational thinking in IR theory. In part two I provide an account of mechanism-based explanation. In part three I deal with each of the three main obstacles

[8] See, for example, Goddard and Nexon, 'Paradigm Lost', pp. 12–13.

[9] See Waltz, *Theory*, ch. 4.

[10] See, for example, Alexander Wendt, 'The Agent-Structure Problem in International Relations Theory', *International Organization*, 41:3 (1987), pp. 335–70; and Wendt, *Social Theory*. See also David Dessler, 'What's at Stake in the Agent-Structure Debate?', *International Organization*, 43:3 (1989), pp. 441–73; Colin Wight, *Agents, Structures and International Relations: Politics as Ontology* (Cambridge and New York: Cambridge University Press, 2006); and Jonathan Joseph and Colin Wight (eds), *Scientific Realism and International Relations* (New York and Basingstoke: Palgrave Macmillan, 2010).

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

mentioned above – context, multiple realisability, and emergence. Let me be clear from the outset that while the scope of this article is relatively broad, it is far from comprehensive, and a number of issues are left out or addressed only briefly. The aim is not to deal with everything, but to provide a starting point for further exploration of the microfoundational strategy.[11]

The search for microfoundations in world politics

IR theory is the scene of some of the most macroscopic research found in social science. In some work, the entire world system is treated as an explanatory variable. IR theory is not, of course, unique in this respect. If we think of the lowest level of nature being that of particle physics, then all sciences engage in macro-explanations of some sort or another – physics included. Macro-explanations provide us with a good deal of understanding, and we should be happy with good explanations wherever we find them. But often, physical and social scientists are not content with macro-explanations on their own. Many scientists choose to look inside these 'black boxes' and identify what, at lower levels of analysis, is doing the real causal work – to identify mechanisms, or microfoundations.[12]

The same tendency is found in IR theory. As is often observed, micro-level research is dominant. Much of it seeks to explain how collective phenomena like institutions, regimes, and anarchy emerge from the interactions of agents. Neo-functionalism, neoliberal institutionalism, constructivism, and Moravcsik's new liberalism are all (in spite of their differences) examples of this tendency.[13] Even better examples, in my view, are found in systemic IR theory. As noted above, systemic theorists usually support their structural explanations with at least some account of underlying mechanisms.[14] This is found even in the work of system theory's staunchest proponents. Waltz is a prime example. While critical of reductionist theory, he is careful not to isolate his own theory from micro-level affairs, and actually brings in the latter to help explain how the system exercises its influence. To explain how anarchy produces patterns of balancing behaviour, for example, he brings in interaction-level processes of 'competition' and 'socialisation'.[15] To explain how polarity affects the system's

[11] One particularly important issue I do not tackle is the meaning of the term 'structure'. Here I use it in a relatively broad and conventional sense, to refer to both social entities (for example, institutions, societies, systems) and their properties (for example, unstable, anarchic, bipolar) – the kinds of things mechanism-based explanations normally target, and which reductionists and holists disagree most sharply about. Alternative uses of the term can be found in some work on the agent-structure problem.

[12] Jon Elster, *Explaining Technical Change* (Cambridge: Cambridge University Press), pp. 23–4; and Elster, *Nuts and Bolts*, pp. 3–10.

[13] See, for example, Ernst B. Haas, *Beyond the Nation State: Functionalism and International Organization* (Stanford: Stanford University Press, 1964); Robert O. Keohane, *After Hegemony: Cooperation and Discord in the World Political Economy* (Princeton: Princeton University Press, 1984); Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984); Checkel, 'The Constructivist Turn'; and Moravcsik, 'Taking Preferences Seriously'.

[14] I have in mind, in addition to the examples I discuss below, virtually all that have engaged in some kind of systemic theorising, from Morton Kaplan to Immanuel Wallerstein to John Mearsheimer. Few IR theorists are strict systemic theorists, but this is part of the point I want to make – they recognise the explanatory value of lower levels of analysis, and the need for microfoundations of some kind. For overviews of various parts of the literature, see Robert Jervis, *System Effects: Complexity in Political and Social Life* (Princeton: Princeton University Press, 1997); and Jack S. Levy and William R. Thompson, *Causes of War* (Chichester: Wiley-Blackwell, 2010).

[15] Waltz, *Theory*, pp. 73–7.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

stability, he describes the decision-making conditions that multipolarity and bipolarity make for states, and explains why decisions to go to war and not go to war are rational in each context.[16] To describe the structure of the system he talks about both the character of the units, and about the distribution of capabilities amongst them.[17] To explain how the system got the way it is (or was), he brings in classical economic theory on market formation – an unabashedly individualistic theory.[18] Waltz is not content just to state a series of systemic generalisations, even though this might have sufficed for a purely structural theory. To be sure, his theory is far from reductionist. He insists that the causal roles structure plays cannot, ultimately, be explained in terms of the system's units. But he does use the units to help make better sense of the system.

Singer had a similar approach. He made extensive use of systemic variables in his research, but was never satisfied just with finding correlations. He argued for an explicitly 'aggregative' view of the system's structural and cultural properties, according to which they are grounded in and can be inferred from the properties of the system's units.[19] In taking this view, he wanted to remove the non-scientific 'mysticism' surrounding system-level phenomena. He did not advocate microfoundationalism *per se*, but did think there would be descriptive and explanatory gains in analysing the system's components.[20] This reductive view of systemic variables has since been echoed by a number of theorists of war.[21]

In general, research on war treats correlations between system-level variables and war outcomes not as complete explanations in themselves, but as guidelines for speculation about underlying causes. Research on polarity and polarisation is a prime example, as noted in the introductory section. Where these are found to influence war and system stability, researchers often try to explain the correlations by describing the local, tangible decision-making conditions that system properties create for states (or leaders), and/or speculating about why states (or leaders) would respond to these conditions in the way that the correlations suggest – often employing decision models in the process. With these 'explanations of explanations', theorists of war try to remove some of the mystery surrounding their correlations. The disputed effects of bipolarity and multipolarity on system stability, for example, are often explained with reference to the conditions of lesser and greater uncertainty they create for states, as Bueno de Mesquita shows.[22] It is generally thought that bipolarity reduces uncertainty, while multipolarity increases it. Accordingly, those that think multipolar systems are more war-prone (and bipolar systems more stable), like Waltz, justify this by arguing that states are more 'risk-acceptant' under uncertainty –

[16] Ibid., ch. 8. See also Kenneth Waltz, 'The Stability of a Bipolar World', *Daedalus*, 93:3 (1964), pp. 881–909.
[17] Waltz, *Theory*, ch. 5.
[18] Ibid., pp. 88–93.
[19] See, for example, J. David Singer, 'Man and World Politics: The Psycho-Cultural Interface', *Journal of Social Issues*, 24:3 (1968), pp. 140–4; and 'Accounting for International War: The State of the Discipline', *Journal of Peace Research*, 18:1 (1981), p. 5.
[20] Singer, 'Man and World Politics', p. 143.
[21] See, for example, Randolph M. Siverson and Michael P. Sullivan, 'The Distribution of Power and the Onset of War', *The Journal of Conflict Resolution*, 27:3 (1983), p. 474; and Bruce Bueno de Mesquita and David Lalman, 'Empirical Support for Systemic and Dyadic Explanations of International Conflict', *World Politics*, 41:1 (1988), p. 1 and 14.
[22] See Bueno de Mesquita, 'Systemic Polarization'; and 'Neorealism's Logic and Evidence: When is a Theory Falsified?', in John A. Vasquez and Colin Elman (eds), *Realism and The Balancing of Power: A New Debate*, (Upper Saddle River, NJ: Prentice Hall, 2003), pp. 166–99.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

given that uncertainty makes going to war a more risky endeavour.[23] Those that think multipolar systems are more stable, on the other hand, argue that states are cautious and 'risk-avoidant' under uncertainty.[24] Alliance polarisation, for its part, is thought to amplify the spread of wars. Polarisation is usually defined as the extent to which alliances cluster around two poles. Its war-expanding role is typically accounted for with reference to the alliances that underlie it. Alliances are, as Vasquez puts it, a 'contagion mechanism' – they spread wars. As such, they 'actualize' polarisation's war-spreading 'potential'.[25]

Much the same inclination is found in constructivist work at the system level. Wendt's work is a prime example. A self-professed holist, he even criticises traditional accounts of structure as capability-distributions, like Waltz's, for being too reductive. But he is careful not to divorce structure entirely from its agent-level bases, using agency to help explain how structure is able to do the work he attributes it in the first place. In *Social Theory of International Politics*, for example, Wendt accounts for the emergence and change of structures of collective knowledge (that is, cultures of anarchy) with reference to unit-level processes like symbolic interaction and socialisation.[26] He accounts for the persistence of structures with reference to the common belief-sets that underlie and constitute them.[27] And he accounts for structures' behaviour-influencing roles with reference to the properties of agents they constitute.[28] In Wendt's theory, structure exercises its influence primarily by means of its constitutive relation to agency. As he puts it, structures 'acquire their causal efficacy only through the medium of practical consciousness and action'.[29] Wendt is using agent-level phenomena to make sense of causal work he otherwise attributes to structure. It might even be argued that Wendt is piggy-backing structure in via an essentially agent-level theory.[30] In any case, it is clear that Wendt would not be satisfied with a purely structural theory of IR. He, like Waltz and virtually everyone else, does not want to make a black box out of structure.

The microfoundational strategy

These are not examples of full-fledged microfoundational explanations by any means. Most leave aspects of structural forces unexplained, and are not couched explicitly in terms of mechanisms or microfoundations. But what is microfoundationalism anyway, and what would a 'full-fledged' mechanism-based explanation look like?

Mechanism-based explanation: a primer

Microfoundationalism – as Wendt and Leon refer to it[31] – has its roots in methodological individualism. This well-known doctrine is, as Lukes puts it, 'a prescription

[23] See Waltz, 'Stability of a Bipolar World'; and *Theory*, ch. 8.
[24] See, for example, Deutsch and Singer, 'Multipolar Power Systems'.
[25] Vasquez, *War Puzzle Revisited*, pp. 269, 325.
[26] Wendt, *Social Theory*, ch. 7.
[27] Ibid., ch. 4.
[28] Ibid.
[29] Wendt, 'Agent-Structure Problem', p. 359.
[30] Indeed, this criticism has already been made of structuration theory, on which Wendt partly bases his own. See, for example, Walter Carlsnaes, 'The Agency-Structure Problem in Foreign Policy Analysis', *International Studies Quarterly*, 36:3 (1992), pp. 245–70; and Wight, *Agents, Structures*.
[31] See Wendt, *Social Theory*, pp. 152–4; and Leon, 'Reduction'.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

for explanation', asserting that all explanations of social phenomena should be 'couched wholly in terms of facts about individuals'.[32] Taken to its logical extreme, it seems to demand that all theories of social phenomena be reduced to individualistic ones. This has been found to be highly problematic, however,[33] and many if not most individualistically inclined social scientists now accept the more sober but no less interesting view that, as Wendt puts it, we endeavour to 'identify the micro-level *mechanisms* by which macro-structures achieve their effects'.[34] Debate now turns largely on the issue of whether mechanisms and microfoundations can be provided for structural phenomena.[35]

Mechanism-based explanation is the means by which microfoundations are provided, and is the core of the microfoundational approach. Also referred to as 'mechanistic' or 'reductive' explanation, mechanism-based explanation has become something of a hot topic in biology, philosophy of science, and social science, thanks largely to the spread of the mechanistic worldview, mentioned earlier. According to this view, causes are mechanisms, and explanation is a matter of uncovering them. The most elaborate accounts of mechanism-based explanation are to be found in philosophy, especially the philosophy of biology and the philosophy of mind. I draw on each in the following account.

Mechanism-based explanations come in two main forms: causal and non-causal.[36] Causal mechanistic explanations, like causal explanations in general, involve specifying the prior conditions responsible for an effect. They show how effects are brought about; they explain changes. The difference between this and more traditional models of causal explanation like deductive-nomological explanation is that the mechanistic approach incorporates a mechanistic notion of causation and causal relevance, which requires looking more deeply into the nature of posited causes (among other things).

More relevant to us, however, is the non-causal form of mechanistic explanation. This is the kind most philosophical discussions focus on. (In the remainder, I refer to it simply as mechanism-based explanation.) This kind of explanation is generally understood as *a process of specifying the mechanisms responsible for causal work associated with higher level phenomena*. It has two especially important features. The first is that it is inter-level, or as some put it, 'componential'. It involves looking inside systems, to their parts, for explanatory mechanisms. As Elster puts it, to explain via mechanisms is to 'open up the black box and show the nuts and bolts, the cogs and wheels of the internal machinery'.[37] Or as Kim puts it, we look to '*explanatory resources at a different level*, a "lower" level in relation to the level of the phenomenon

[32] Steven Lukes, 'Methodological Individualism Reconsidered', *The British Journal of Sociology*, 19:2 (1968), p. 121. See also Lars Udehn, *Methodological Individualism: Background, History and Meaning* (London: Routledge, 2001), p. 354.

[33] On this issue see Harold Kincaid, 'Reduction, Explanation, and Individualism', *Philosophy of Science*, 53:4 (1986), pp. 492–513; and William McGinley, 'Reduction in Sociology', *Philosophy of the Social Sciences*, 42:3 (2012), pp. 370–98.

[34] Wendt, *Social Theory*, p. 153.

[35] See, for example, Hedström and Swedberg, *Social Mechanisms*; and Hedström and Ylikoski, 'Causal Mechanisms' (see fns 3 and 4 above).

[36] On the distinction, see Carl F. Craver, *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience* (Oxford: Oxford University Press, 2007), pp. 107–8.

[37] Elster, *Explaining Technical Change*, p. 24.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

to be explained'.[38] Its other notable feature is that it is non-causal. In a causal explanation (mechanistic or otherwise), the aim is, as Cummins puts it, to 'explain changes of state in a system as effects of previous causes'.[39] In a mechanism-based explanation, however, the phenomenon to be explained is the *behaviour* of a macro-level system. Mechanism-based explanation involves identifying the micro-level conditions that make it possible for a structure to play a certain role. Such explanations are answers to 'how-possible' questions of the form: 'How is it possible that structural factor x plays role y?'

In order to see exactly why this is not causal, we need to know more about the terms 'role' and 'mechanism' here. By role I mean disposition. Mechanism-based explanations usually target general dispositions of systems, rather than isolated behaviours. A disposition is a causal tendency – something an object tends to do, in certain conditions. Dispositions are often referred to as dispositional properties or 'powers'. Classic examples are solubility and flexibility, and mental properties like pain and hunger are also defined this way. Dispositional properties have the special characteristic of being *realised* by properties at lower levels of analysis. The latter are often referred to as realisation 'bases'. Pain for example, is something like the property of tending to wince and retract, in response to tissue damage, and it is realised by physical-chemical states of the brain.[40] The main aim of a mechanism-based explanation is to show how a dispositional property is realised at the micro-level, or identify its realisation base. Realisation-bases are mechanisms. The literature contains a large number of definitions of mechanism, but there are important points of agreement,[41] including that mechanisms are realisers.[42] Craver's definition is a good starting point: a mechanism, he says, is 'a set of entities and activities organised such that they exhibit the phenomenon to be explained'.[43] The 'phenomenon to be explained' is a causal role played by a higher level property; the mechanism is what realises this property and its role. Realisation relations are not causal – realisation-bases and dispositions instantiate instantaneously.[44] Pain, for example, is a brain-state usually caused by tissue damage, but realised by a person's nervous system. A causal explanation of a person's pain would cite the events that preceded and led to the tissue-damage, and the tissue-damage that in turn led to the pain. A mechanism-based explanation, in contrast, would talk about the state of the person's nervous system *at the time of* the pain.

The examples I cited in the previous section were all at least partial examples of explanation of this kind. Alliance polarisation's war-spreading role, for example, was

[38] Jaegwon Kim, *Physicalism, or Something Near Enough* (Princeton: Princeton University Press, 2005), p. 106. Note that causal mechanistic explanations can also be inter-level, relating lower-level causes to higher-level effects.

[39] Robert Cummins, *The Nature of Psychological Explanation* (Cambridge, MA: MIT Press, 1983), p. 1.

[40] The description is Kim's. See Jaegwon Kim, *Philosophy of Mind*, second edn (Boulder, CO: Westview Press, 2006).

[41] See the reviews of the literature provided in Craver, *Explaining the Brain*; and Hedström and Ylikoski, 'Causal Mechanisms'.

[42] See, among others, Carl Gillett, 'Understanding the New Reductionism: The Metaphysics of Science and Compositional Reduction', *The Journal of Philosophy*, 104:4 (2007), pp. 193–216; Craver, *Explaining the Brain*; and Thomas W. Polger, 'Mechanisms and Explanatory Realization Relations', *Synthese*, 177:2 (2010), pp. 193–212.

[43] Craver, *Explaining the Brain*, p. 5.

[44] See Thomas W. Polger, 'Realization and the Metaphysics of Mind', *Australasian Journal of Philosophy*, 85:2 (2007), pp. 233–59. Philosophers disagree about the exact nature of realisation relations, a matter I will not get into here. They do, however, agree that they are non-causal.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

explained in terms of unit-level conditions – alliance linkages. Polarity's stability-influencing role was explained with reference to conditions of uncertainty and utility calculations of states. In these and the other cases, the object of explanation is a causal role played by a system property, and the entities responsible constitute part of their realisation bases. Consider the explanation of polarisation. Alliance linkages do not cause alliance polarisation to spread wars. They constitute alliance polarisation, and actually *fulfil* this role themselves. This is an explanation of alliance polarisation's causal capacity, which is a property it possesses, rather than an effect or outcome.[45]

Perhaps surprisingly, this sort of explanation is not commonly found in reductionist IR research. Reductionist theories are generally seen as those that explain international outcomes in terms of micro-level causes, like the attributes and interactions of states, or the attributes and interactions of decision-makers within states.[46] Some go a step further and try to explain the emergence of institutions or regimes from the interactions of actors, as noted above. These sorts of theories are reductive in the traditional sense of exhibiting a micro-level explanatory orientation – they invoke micro-level causes for macro-level effects. But they do not, for the most part, provide many examples of mechanism-based explanation *proper*, because they are primarily causal. They treat institutions as outcomes, and explain them by describing the processes of interaction that *cause* them to emerge. They are, as Cummins would say, transition theories – they explain the system's transition from lacking a certain institution, to gaining it; or from having a regime of one kind, to having one of another kind.[47] Still, this sort of work may be regarded as being microfoundational in a less direct sense. These theories suggest that lower levels of analysis are where most of the real causal work gets done in world politics. This, in turn, seems to imply that how things are at lower levels determines how things are at higher levels. If systemic factors appear to being doing any causal work, then it is probable that this is really being fulfilled at lower levels. What are lacking in these theories, however, are explanations linking system-level causal roles and micro-level mechanisms directly.

The distinction is crucial, because each sort of explanation provides us with a fundamentally different kind of explanatory information. Causal explanations tell us how outcomes or changes are brought about; mechanism-based explanations tell us how the causes cited in causal explanations actually do the causal work assigned to them. Causal explanations attribute causal roles to things, in order to explain why outcomes or changes occur. In effect, they attribute dispositions to them. Those that explain the spread of war in certain periods with reference to alliance polarisation attribute it, in effect, a war-spreading disposition. This often prompts an additional question, as to how it is possible that the causal entity or property in question is able to do the causal work assigned to it. In virtue of what does alliance polarisation spread wars? How is it able to fulfil this role? The answer is not some prior event, but mechanisms found in its own realisation-base.

[45] Mechanism-based explanation resembles what Wendt has called 'constitutive explanation'. The latter involves identifying the conditions that make it possible for an object or event to have a certain property. Wendt mentions constitutive explanations of dispositional properties, but focuses more on explanations of qualitative properties – properties of 'being' this or that, rather than 'doing' this or that. See Alexander Wendt, 'On Constitution and Causation in International Relations', *Review of International Studies*, 24:5 (1998), pp. 101–17.

[46] Waltz, *Theory*, p. 61. See also James D. Fearon, 'Domestic Politics, Foreign Policy, and Theories of International Relations', *Annual Review of Political Science*, 1 (1998), pp. 289–313.

[47] These theories might be interpreted as providing mechanistic explanations of a causal variety instead, but this is beyond the scope of this article.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

Why is it that mechanism-based explanations clarify macro-level explanations, and not the other way round? This has the appearance of being arbitrary. But from the point of view of microfoundationalism, it is quite reasonable. This 'downward' orientation is motivated primarily by the idea that lower levels of organisation are more causally fundamental than higher ones. All things have parts, and it is natural to think that these parts are a necessary condition for their existence and for any causal work they do. Philosophers often articulate this point with reference to ideas like physicalism, 'supervenience', and ontological individualism. Physicalism is the view that there is only one substance in the universe – matter – and all things and processes are essentially physical. Ontological individualism is physicalism's sociological cousin, and says that there is nothing more to social life than individual-level stuff – individuals, their properties, interactions, and so on. Both are often stated in terms of supervenience. Supervenience is a dependence relation between properties of objects – property *a* supervenes on property *b* if *a* could not differ without *b* also differing.[48] The mind supervenes on the brain, in the sense that people in identical brain states will be in identical mental states. Society supervenes upon individuals in the sense that, as Sawyer puts it, 'if a collection of individual properties with a given set of relations causes a certain social property to obtain on one occasion, then that same collection of individual properties in that same set of relations on another occasion will cause the same social property to obtain'.[49] Physicalism and supervenience are both popular theses in philosophy,[50] and ontological individualism is also well-regarded, especially when stated as a supervenience thesis.[51] Together they have the implication that how things are at lower levels determines how things are at higher levels. The individual level carries, or does the causal work of, the macro-level.[52] This is why, at least for the microfoundationalist, it is rational and not at all arbitrary that we should look to lower levels for clarification and explanation of macro-level phenomena, and not the other way around. These ideas are not without their detractors, however, and I consider one prominent criticism – emergence – below.

A three-step strategy

Thus far I have described what mechanism-based explanations are and tend to look like. I have not yet said how they are built. A number of elaborate strategies for producing mechanism-based explanations are to be found in the literature. Some particularly interesting and useful ones are Kauffman's 'articulation of parts explanation', Cummins's 'functional analysis', Bechtel and Richardson's 'decomposition and localization' heuristic, and Kim's 'functional reduction'.[53] Each involves at least three steps: (i) *functional analysis*, (ii) *mechanism search*; and (iii) *reductive explanation*.

[48] See Jaegwon Kim, *Supervenience and Mind* (New York: Cambridge University Press, 1993).
[49] R. Keith Sawyer, 'Nonreductive Individualism: Part I – Supervenience and Wild Disjunction', *Philosophy of the Social Sciences*, 32:4 (2002), p. 543.
[50] See Kim, *Philosophy of Mind*, pp. 8–14 and 273–5.
[51] See Brian Epstein, 'Ontological Individualism Reconsidered', *Synthese*, 166:1 (2009), p. 188.
[52] See, for example, Elster, *Explaining Technical Change*, pp. 28–9; and Frank Jackson and Philip Pettit, 'In Defense of Explanatory Ecumenism', *Economics & Philosophy*, 8:1 (1992), p. 5.
[53] See, respectively, Stuart A. Kauffman, 'Articulation of Parts Explanation in Biology and the Rational Search for Them', *Boston Studies in the Philosophy of Science*, 8 (1970), pp. 257–72; Cummins, *Psychological Explanation*; William Bechtel and Robert C. Richardson, *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research* (2nd edn, Cambridge, MA: MIT Press, 2010); and Kim, *Physicalism*.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

The first step, functional analysis, involves identifying the causal role of the system or structure to-be-explained. This is to be treated as the *explanandum*-phenomenon. The second step involves a search for the mechanism or mechanisms that realise the causal role in question. One looks amongst the system's components, their properties, interactions, and organisation, with the aim of finding conditions sufficient to realise the causal role in question. The final step involves producing the explanation. One constructs a general theory or model to explain the correlation between the realising mechanisms and the realised causal role. The theory or model could be constructed within the apparatus of a relevant existing micro-level theory. The second step – mechanism search – is the most important, however, and the one most accounts focus on. Bechtel and Richardson, in their now-classic work on the matter, propose a strategy of decomposition and localisation, or 'hierarchical analysis into functional components with specific functions'.[54] Decomposition involves subdividing the overall function or causal role into a series of sub-tasks or 'set of subordinate functions' that together comprise 'the behaviour we are studying'.[55] Localisation involves associating each sub-task with a specific component of the system.[56]

This strategy is especially useful in cases of multiple realisability, as I will argue below. Strategies like it have been discussed in IR theory before. A good example is Most and Starr's 'nice law' strategy.[57] They premise this strategy on the observation that certain consistent foreign policy outcomes can have 'alternative triggers' at the decision-making level. Self-help behaviour, for example, can be pursued on the basis of a variety of goals, even while external constraints like anarchy and the distribution of power remain constant. Thus, system-level generalisations about self-help behaviour (like Waltz's) capture necessary, but not sufficient conditions for self-help behavioural tendencies.[58] Most and Starr argue that in order to account for alternative triggers, we break down system-level generalisations into sets of nice, 'sometimes true', or 'domain-specific' generalisations, each of which identifies a different kind of motivation for the behaviour in question. This can be interpreted as involving decomposition and localisation. In their revision of the neorealist self-help generalisation, for example, Most and Starr decompose system structure's role of inducing self-help behaviour into a set of specific sub-tasks (inducing militaristic self-help behaviour, inducing domestic-political self-help behaviour, inducing economic self-help behaviour, etc.), and localise each function to particular sorts of goals (security goals, political goals, economic goals, etc.).[59]

Mechanism-based explanation and micro-macro problems

As noted above, the primary function of a mechanism-based explanation is to account for a structural factor's causal powers. In this way mechanism-based explana-

[54] Bechtel and Richardson, *Discovering Complexity*, p. 7.
[55] Ibid., pp. 23–4.
[56] Ibid. See Craver, *Explaining the Brain*, for a different, counterfactual approach.
[57] See Benjamin A. Most and Harvey Starr, 'International Relations Theory, Foreign Policy Substitutability, and "Nice" Laws', *World Politics*, 36:3 (1984), pp. 383–406.
[58] Ibid., pp. 405–6.
[59] This is only a partial example of a microfoundational explanation, however, because in their theory system structure still plays a constraining, opportunity-determining role. Keohane's 'issue-area' disaggregation strategy also seems to be in the spirit of mechanism-based explanation, and localisation in particular. See Robert O. Keohane, 'Theory of World Politics: Structural Realism and Beyond', in Robert O. Keohane (ed.), *Neorealism and its Critics* (New York: Columbia University Press, 1986), pp. 182–90.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

tions add depth to causal explanations.[60] Another way of putting it is that mechanism-based explanations are 'mystery-removing' explanations.[61] Causal explanations remove mysteries of their own, concerning why things happen. But they inadvertently create new mysteries as well, concerning how causal factors are able to do their causal work in the first place. Mechanism-based explanations address the latter sort of mystery. They do this by providing microfoundations – a system's microfoundations being its realising mechanisms. This is particularly attractive from the point of view of scientific realism, for which explanation is about getting to the underlying essences of things, to fully account for their causal powers.[62]

Because it involves linking higher- and lower-level phenomena, some philosophers also see mechanism-based explanation as a means of resolving micro-macro problems. Some argue it might be able to resolve the mind-body problem in philosophy,[63] and others have suggested it could do the same for the holism-individualism debate in social science.[64] In my view – and whatever its results in these areas – the microfoundational strategy might help resolve IR theory's own micro-macro problems. Two problems of this kind have been discussed at length in IR theory: the level-of-analysis problem and the agent-structure problem. The former, it is said, concerns the relative explanatory value of different levels of analysis in world politics – levels usually being seen as a system for organising variables, based mainly on their scale.[65] The microfoundational strategy appears to favour lower levels, because of its suggestion that lower levels are more causally fundamental. But the strategy is actually as conducive to macro-level analysis as it is to micro-level analysis. By removing mystery from structure – a macro-level entity – mechanism-based explanations should *enhance* its credibility as an explanatory tool. Indeed, this seems to be the intention behind the mechanism-based explaining that already goes on in systemic IR theory. Why would microfoundations enhance structure's credibility? For at least two reasons. First, because lower levels of organisation are more causally fundamental, at least according to microfoundationalism. Going down in level gets you closer to physical mechanisms. It helps us integrate higher level processes into an underlying ontology. The second reason is that microfoundations help bring structural theories into closer alignment with theories at lower levels. A theory's consistency with the rest of science, especially with the best-established sciences, is an important measure of its credibility.[66] Thus, in my view, and counter to initial appearances, mechanism-based explanation should be most attractive to systemic theorists of IR.

Microfoundations are, of course, good for reductionists too. By showing that higher level processes can be accounted for with lower level explanatory resources, more *explananda*-phenomena are brought within the explanatory grasp of reductionist

[60] Hedström and Ylikoski, 'Causal Mechanisms', p. 54.
[61] David J. Chalmers, *The Conscious Mind* (New York: Oxford University Press, 1996), p. 48.
[62] See, for example, Rom Harré and Edward H. Madden, *Causal Powers: A Theory of Natural Necessity* (Totowa, NJ: Rowman and Littlefield, 1975); and Wesley C. Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton: Princeton University Press, 1984).
[63] See Kim, *Physicalism*.
[64] See McGinley, 'Reduction'.
[65] See, for example, J. David Singer, 'The Level-of-Analysis Problem in International Relations', *World Politics*, 14:1 (1961), pp. 77–92; and Hans Mouritzen, 'Selecting Explanatory Level in International Politics: Evaluating a Set of Criteria', *Cooperation and Conflict*, 15:3 (1980), pp. 169–82.
[66] See Patricia Smith Churchland, *Neurophilosophy: Toward a Unified Science of the Mind-Brain* (Cambridge, MA: MIT Press, 1986), p. 376.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

theory, and its 'leverage' is increased.[67] Mechanism-based explanations also reveal as much about micro-level mechanisms as they do structural phenomena. Primarily, they tell us more about what lower level mechanisms *do*. Without explanations linking mechanisms with higher level roles, we would only know about the roles mechanisms play at their own level. Structural theories, via mechanism-based explanations, help microfoundationalists see the bigger picture – see the functional forest from the mechanical trees, so to speak. The explanation of alliance polarisation in terms of alliance commitments, for example, helps us see that alliances can spread wars throughout systems, not just within dyads – despite the fact that they are normally intended to mitigate war. And Waltz's explanation of anarchy's selection-effects in terms of competition and socialisation helps us see that the latter two mechanisms play a role in generating system-wide patterns of balancing behaviour.

Microfoundationalism's solution to the level-of-analysis problem is, then, relatively pluralistic – no particular level gets priority. In itself this is not a novel solution, nor is it likely to be met with much opposition, for most that have written on the level-of-analysis problem already recognise that no single level – micro or macro – is capable of doing all the explanatory work required of IR theory.[68] Microfoundationalism is not so neutral when it comes to the agent-structure problem, however. This problem, it is said, concerns how we conceptualise the relation between agents and structures in world politics.[69] Mechanistic explanation is all about capturing micro-macro relations like this. More specifically, it is designed to capture relations between macro-level phenomena and their realising mechanisms. Where successful, mechanism-based explanations show that structural phenomena are wholly dependent for their causal powers on agent-level mechanisms. Microfoundationalism's proposed solution is that we close the gap between agency and structure, as it were, by means of mechanism-based explanations. The agent-structure relation, on this view, is one of asymmetric dependence – structures depend on agents, at least for their causal potentialities.

This contrasts sharply with the way most writing on the agent-structure problem have depicted the agent-structure relation. Most argue that neither structure nor agency is primary, and that each plays a part in determining the other, in a kind of recursive interplay.[70] These authors are, as it is sometimes put, dualists – other differences in their views notwithstanding. Some adopt the criticisms I review below to undermine the microfoundational view and support their dualism. The success of mechanism-based explanation as a solution to the agent-structure problem turns in large part on these criticisms failing, a matter to which I now turn.

[67] On explanatory leverage, see Gary King, Robert O. Keohane, and Sidney Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton: Princeton University Press, 1993), pp. 29–31.

[68] See, for example (and among several others), Singer, 'Level-of-Analysis Problem'; J. David Singer, 'From A Study of War to Peace Research: Some Criteria and Strategies', *The Journal of Conflict Resolution*, 14:4 (1970), p. 537; Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976); Peter Gourevitch, 'The Second Image Reversed: The International Sources of Domestic Politics', *International Organization*, 32:4 (1978), pp. 881–912; Mouritzen, 'Selecting Explanatory Level'; Most and Starr, 'Foreign Policy Substitutability'; and Levy and Thompson, *Causes of War*. I say more on this below.

[69] See, for example, Wendt, 'Agent-Structure Problem'.

[70] See ibid.; as well as Dessler, 'What's At Stake'; Carlsnaes, 'Agency-Structure Problem'; and Wight, *Agents, Structures*, among several others.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

Obstacles to the microfoundational strategy

Mechanism-based explanation, I have tried to show, is an elaborate strategy that should appeal to reductionists and structuralists alike. In practice, however, the strategy faces some significant obstacles. It has been attacked on multiple fronts, most notably by Waltz and Wendt. I mentioned three criticisms in the introductory section: the role of context, multiple realisability, and emergence. Each appears to make mechanism-based explanations of structural phenomena unfeasible. I will argue, however, that when properly understood, the microfoundational strategy should be able to overcome each of these obstacles. Space constrains me from dealing with the obstacles as fully as they deserve, but I can provide starting points. There are other obstacles too of course, but these three appear to be the most pressing.

Obstacle 1: structure and context

The first and most apparent obstacle is structure itself. Microfoundationalism is a reductive research strategy. In IR theory, reductionism is usually associated with a domestic, 'inside-out' explanatory orientation on world politics. This is evident in meta-theoretical discussions of reductionist IR theory and of the level-of-analysis problem.[71] Since mechanism-based explanation involves using the explanatory resources of lower levels of analysis, it would seem to demand a domestic, inside-out view of world politics as well. To explain system-level phenomena, microfoundationalism would have us look to the attributes and interactions of states, or perhaps even further inside states, to bureaucratic structures or decision-makers. Microfoundationalism, then, seems likely to marginalise structure. This would be a problem. While systemic views of world politics are not especially popular, most concede that structure plays some role in world politics. As noted above, most that have written on the level-of-analysis problem recognise the explanatory value of multiple levels. As Levy and Thompson observe, most theories of IR incorporate variables from multiple levels of analysis.[72] Even those that declare a preference for lower levels of explanation, like liberal and decision-making theorists of foreign policy, still maintain that the systemic environment plays some kind of role – usually that of a distal constraint on choice or opportunity.[73] Theories now generally transcend particular levels.[74] This has lead some working on the agent-structure problem to declare that

[71] See, for example, Singer, 'Level-of-Analysis Problem'; Jervis, *Perception*; Waltz, *Theory*; Gourevitch, 'Second Image Reversed'; Most and Starr, 'Foreign Policy Substitutability'; and Levy and Thompson, *Causes of War*. The view of the debate as a simple reductionist/systemic dichotomy is long gone of course, but this does not mean the terms no longer apply – they just apply in a more complex, relativistic way. See Martin Hollis and Steve Smith, *Explaining and Understanding International Relations* (Oxford: Clarendon Press, 1990); and Fearon, 'Domestic Politics'.

[72] Levy and Thompson, *Causes of War*, pp. 18, 207.

[73] See, for example, Jervis, *Perception*; Most and Starr, 'Foreign Policy Substitutability'; Moravcsik, 'Taking Preferences Seriously'; and James M. Goldgeier and Philip E. Tetlock, 'Psychology and International Relations Theory', *Annual Review of Political Science*, 4 (2001), pp. 67–92.

[74] See, for example, Robert D. Putnam, 'Diplomacy and Domestic Politics: The Logic of Two-Level Games', *International Organization*, 42:3 (1988), pp. 427–60; Fearon, 'Domestic Politics'; and Suzanne Werner, David Davis, and Bruce Bueno de Mesquita, 'Dissolving Boundaries: Introduction', *International Studies Review*, 5:4 (2003), pp. 1–7.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

structure's role in political life is 'uncontentious'[75] and even a 'truism',[76] and dismiss individualistic views like microfoundationalism offhand.[77] Waltz and Singer have dismissed reductionism on similar grounds.[78]

That the systemic environment plays a causal role in the decision-making of states and leaders seems undeniable. How, then, could the microfoundational strategy possibly work? Fortunately, the microfoundational strategy is not restricted to a purely domestic orientation. The aim with this strategy, I have been stressing, is to account for the behaviours or dispositions of systems with reference to their components and properties of their components. *Context itself is the explanandum-phenomenon*. Furthermore, context is part of the *explanans* as well. Consider the examples of section one again. Alliance polarisation's war-spreading role was explained not in terms of any domestic factor, but alliances *between* states. Polarity's stability-influencing roles were explained partly with reference to the dispositions of states, but also with reference to environmental uncertainty. These explanations refer us to the microfoundations of context itself, to explain how its influence is realised. They exhibit a decomposition and localisation strategy. Some of these microfoundations (for example, attributes, dispositions) are domestic in nature, but some (for example, relations, conditions) are not. The latter are contextual in nature, but this does not mean they are automatically systemic. Context is not essentially a macro-social thing. Like all substantial things, context is built out of parts, and we can break it down into parts of virtually any scale. The microfoundational strategy involves doing exactly this.

Microfoundationalism is also conducive to the use of large-scale combinations of micro-level variables. Given its large scale, the systemic context probably does most of its causal work in large spatial and temporal sweeps, producing system-wide and period-spanning outcome patterns. Rather than particular foreign policy choices (that is, particular wars or cooperative ventures) it more likely affects things like overall system stability.[79] Mechanism-based explanations of systemic tendencies will not get far if the focus is just on isolated, small-scale, domestic variables.[80] These rarely suffice to realise whole systemic dispositions.[81] Mechanism-based explanation would instead require showing how the parts (in plural) of the system interact, work together, and are organised, so as to realise its causal roles. Many of these parts might be, individually speaking, domestic in nature. But clustered together, they turn into large scale variables, and ones that are contextual, relative to individual states.

One might wonder why, if mechanism-based explanations can reference large-scale factors like context and structure, they can be considered 'reductive' or 'micro-foundational'. There are actually good reasons for thinking so, and they can be brought out by looking at precedents in social science and philosophy. Consider the

[75] Dessler, 'What's At Stake', p. 443.
[76] Wendt, 'Agent-Structure Problem', pp. 337–8.
[77] See, for example, Carlsnaes, 'Agency-Structure Problem', pp. 248–50; and Wight, *Agents, Structures*, p. 131.
[78] See Waltz, *Theory*; and Singer, 'From A Study of War to Peace Research', p. 537.
[79] See, for example, Waltz, *Theory*.
[80] Levy and Thompson , *Causes of War*, pp. 18–20, make a similar point.
[81] But they may in some cases. See Siverson and Sullivan, 'Distribution of Power', p. 474; and Bueno de Mesquita and Lalman, 'Empirical Support'.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

mind/body problem in philosophy. This problem concerns whether mental states reduce to physical-chemical states of the brain, and pits reductionists against 'emergentists'. Reductionists do not say that mental states reduce to states of particular particles or microscopic regions of the nervous system, but to states of the whole nervous system – that is, its components in aggregate. This is a genuinely reductionist point of view, as Kim argues, because they are saying mental states are not emergent, but reduce to 'micro-based' properties – clusters of lower-level properties.[82]

Consider also the holism-individualism debate. Individualism does not demand omission of all reference to context in the explanation of social outcomes. The idea that all social outcomes can be explained with reference to factors internal to individuals, like psychological attributes, is generally regarded as utterly implausible. Individualism does not say that context has no explanatory role, just that *macro*-social things have no explanatory role. And context is not essentially a macro-social thing, as we have seen. Features of context that are individualistic in scale – like other agents, relations, actions, interactions, accessible physical objects, resources, and so on – can and do feature in individualistic explanations, and standard definitions of methodological individualism attest to this.[83] Individualists even use variables traditionally associated with holism, like structure and culture, but define them aggregatively. The distinction here is essentially the same as Wendt's distinction between cultures of common knowledge and cultures of collective knowledge.[84] Common knowledge, Wendt says, is merely a common set of beliefs held by individuals, and is the kind of culture microfoundationalists recognise. Neoliberals and many constructivists conceive of culture in this way. Singer defined culture in a very similar way, asserting that 'the cultural properties of any subnational, national, or extranational system may be described in a strictly aggregative fashion, by observing the distribution and configuration of individual psychological properties'.[85] True holists, on the other hand, view cultures as structures of collective knowledge, which while constituted by common beliefs, do not to reduce to them – at least on the holist account.[86]

Overall, two things make mechanism-based explanations of structure and culture reductive. The first is the aggregative view of structure implied. The second is the fact that, in the final analysis, the only mechanisms attributed basic causal power are micro-level ones. In mechanism-based explanations, the parts (and their properties) are doing all the real causal work. In genuinely structural explanations, on the other hand, like those Waltz and Wendt offer, structures are attributed irreducible causal or explanatory power of their own, and depicted in a non-aggregative (that is, emergent) way. The implication is that large-scale contextual variables like culture are not off-limits for the microfoundational approach, as long as the emphasis is always on the parts. This, of course, is generally recognised already. Waltz, for example, says that interactions – which are technically outside agents – are unit level, and

[82] Jaegwon Kim, *Mind in a Physical World* (Cambridge, MA: MIT Press. 1998), pp. 80–7.
[83] Watkins, for example, famously described methodological individualism as saying that every social outcome is a result of a 'particular configuration of individuals, their dispositions, situations, beliefs, and physical resources and environment', and that all social outcomes must be explained in terms of 'the dispositions, beliefs, resources, and inter-relations of individuals'. See Watkins, 'Historical Explanation in the Social Sciences', *The British Journal for the Philosophy of Science*, 8:30 (1957), p. 106.
[84] See Wendt, *Social Theory*, pp. 157–84.
[85] Singer, 'Man and World Politics', pp. 143–4.
[86] Wendt, *Social Theory*, pp. 162–3.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

along with attributes are part of the domain of reductionist theory.[87] Dyad-level and region-level theories of war treat smaller segments of the systemic environment (that is, dyads and regions) as explanatory levels distinct from the system level. There is also recognition of the fact that using large-scale combinations of variables can be reductive. Waltz himself says that reductionist theories 'explain international outcomes through elements and *combinations of elements* located at national or sub-national levels'.[88] Wendt also recognises it, with his distinction between common and collective knowledge.[89]

Structure, then, is not in itself a problem for mechanism-based explanation. Structure is the problem mechanism-based explanation was designed to deal with in the first place. Rather than marginalising structure, as reductionist IR theories sometimes do, mechanism-based explanations would actually bring it back into the fold, albeit in a component-emphasising way.

Obstacle 2: multiple realisability

The real obstacles to microfoundationalism are the next two: multiple realisability and emergence. Each has the implication that structures cannot be explained in terms of realising mechanisms, something that, if true, would scuttle any reductive research strategy. I deal with multiple realisability first. This influential notion is largely responsible for the major backlash against reductionism in recent philosophy, and its effects have been felt all across the sciences.[90] It was only ever a matter of time before it was brought into IR theory. Wendt was the theorist responsible, though he acknowledges that others discussed similar ideas in the past. In *Social Theory*, Wendt argues that system and state-level factors are multiply realisable at lower levels of analysis.[91] He acknowledges, in line with the philosophical argument, the ontologically reductive principle of supervenience, at least for the sake of this argument. As noted above, supervenience suggests that micro-level conditions determine macro-level conditions. But importantly, as Wendt notes, the 'upward determination ... is not strict or 1:1, since it allows for the possibility that the same macro state could be realised by many different micro states' – multiple realisability.[92] This, he argues, means micro-level states of affairs will be '*sufficient* for the existence of a macro-state, but not *necessary*'.[93] Because of this, micro-level inquiry 'may generate *disparate* explanations for events that in fact have a *common* macro-level cause'.[94] He elaborates with the example of a window that breaks when a person (John) throws a rock at it.[95]

[87] See, for example, Waltz, *Theory*, p. 66.
[88] Ibid., p. 60, emphasis added.
[89] See also Wendt's distinction between macro-structure and 'micro-structure'. Wendt, *Social Theory*, pp. 145–56.
[90] For the two most influential statements of the argument, see Hilary Putnam, 'The Nature of Mental States', in Hilary Putnam (ed.), *Mind, Language, and Reality: Philosophical Papers, Volume 2* (Cambridge: Cambridge University Press, 1975), pp. 429–40; and Jerry A. Fodor, 'Special Sciences (Or: The Disunity of Science as a Working Hypothesis)', *Synthese*, 28:2 (1974), pp. 97–115.
[91] Wendt, *Social Theory*, pp. 152–6.
[92] Alexander Wendt, 'The State as Person in International Theory', *Review of International Studies*, 30:2 (2004), p. 300.
[93] Wendt, *Social Theory*, p. 152.
[94] Ibid., p. 154.
[95] Ibid.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

Wendt contends that the best explanation for the broken window is that 'John threw a rock at it', rather than 'an analysis of the particular combination of sub-atomic particles' that broke it. While the exact particle-structure of the rock was causally responsible for the breaking of the glass, 'many other combinations would have had the same effect'. In other words, the macro-cause (John throwing a rock at the window) was multiply realisable at the particle-level. By extension, the macro-effect (the window being broken) could have been brought about in many ways at the particle-level. The only factor stable enough to explain the outcome is, therefore, the macro-cause. Wendt argues that the same is true of systemic tendencies in IR, and concludes that 'the door is open to relatively autonomous macro-level explanations'.[96]

Wendt likens his claim to ones made earlier by Waltz and Most and Starr. Waltz argued that unit-level mechanisms vary more than the international outcomes they are supposed to explain.[97] He concluded, much like Wendt does, that these causes cannot explain these outcomes in an 'elegant' or general way. Instead, systemic variables are needed, because of their relative invariability. Most and Starr made a similar observation to Waltz, but drew a rather different, reductionist conclusion, which I consider shortly. Unfortunately, however, Wendt is not clear what he means by multiple realisability, and it is not immediately clear what the relation between his argument and those of the others is. Sometimes he suggests that multiple realisability is like what Waltz and Most and Starr observe – the capacity of a given macro-level outcome for different micro-level causes – as when he discusses his three cultures of anarchy.[98] Sometimes he suggests it is the capacity of a macro-level entity for different constitutions or arrangements of parts, like when he talks about supervenience.[99] At other points he suggests that it is something like the capacity of a macro-level causal tendency for different micro-level fulfillers, like when he talks about the Waltzian tendency toward 'balancing and institutional isomorphism'.[100]

The correct meaning is actually the third, which is unfortunately the one that Wendt pursues the least. Multiple realisability is *the capacity of a dispositional property for different micro-level realisers*. To realise, as we saw above, means to occupy, fulfil, or implement a causal role.[101] Multiple realisability, in turn, is the capacity of a given causal role for different occupants – same function, different implementing mechanisms.[102] Philosophers illustrate the idea with simple examples like mousetraps. Mousetraps function to catch mice, but the exact mechanisms they use differ greatly. Traditional ones use a spring-loaded bar with a bait-laden trip for release. So-called 'mouth-type' mousetraps use plastic jaws and a coil-spring triggered by a bait-laden mechanism inside. Some use electricity to shock mice. Some trap them in glue. And so on. The function that defines the kind, *mousetrap*, can be realised in very different ways. Multiple realisability is, therefore, neither a causal nor a constitutive relation. Wendt's broken window was caused by the prior event of John throwing a rock at it; the rock's window-breaking capacity was realised by its physical substructure. The potential of a single outcome for different micro-level causes – which is what Waltz

[96] Ibid., p. 156.
[97] See Waltz, *Theory*, pp. 39–40, 65–70, and 77–8.
[98] See Wendt, *Social Theory*, pp. 254, 343, and 364.
[99] Ibid., pp. 155–6; and Wendt, 'State as Person', pp. 299–301.
[100] Wendt, *Social Theory*, pp. 150–2.
[101] See Polger, 'Realization'.
[102] See William G. Lycan, *Consciousness* (Cambridge, MA: MIT Press, 1987), ch. 4.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

and Most and Starr observe – would be better described as something like 'multiple causability'.

When understood in this way, multiple realisability becomes quite an intuitive thesis. Whether we are talking about machines, bodily organs, computer programmes, or psychological states – indeed, anything that can be defined in terms of its causal role – it is natural to think that functions can be realised in different ways. Could the multiple realisability argument-proper be applied to IR theory? I suspect it could, but only if anti-reductionists like Wendt and Waltz are willing to conceptualise structural properties as dispositional properties, realised at the micro-level. Hilary Putnam, with whom the argument originated, used it to show that mental states like pain and hunger are functional states – dispositions of a whole organism, defined in terms of their causal roles, and realised at the physical level. The argument all but entails a functional view of higher order properties like this, something Wendt misses entirely. This is a problem, because Wendt also argues that structures are ontologically emergent from their microfoundations, and this clashes with the physical realisation thesis.

However, even if the argument were successfully applied in IR, there are ways around it. One is contained within the mechanistic approach itself: localisation. In particular, Most and Starr's nice-law strategy, being an example of a localisation strategy, is able to deal with multiple realisability. As noted earlier, Most and Starr argue that, when faced with alternative triggers (that is, multiple causability), the IR theorist should retreat from always-true, system-level generalisations to nice-laws, or domain specific generalisations. From a Waltzian perspective, this seems illogical, because it would mean throwing away perfectly good macro-generalisations. However, in the following passage, Most and Starr make a very interesting point.

> [I]f different processes do indeed lead to similar concrete empirical behaviors, the concept reflected by one such act could be different within different formulations. If the decision makers are dealing with a national viability or security dilemma, for example, an increase in their defense expenditures or the formation of an alliance would be interpretable as – and more importantly, serve as an indicator of – efforts to increase their state's defense capacity or to decrease the foreign risks with which they perceive it to be confronted; if they are seeking to resolve a simple governmental viability problem, on the other hand, those same acts would reflect efforts to increase their government's strength or to decrease the degree to which it is threatened by domestic forces.[103] In other words, it may actually be the case that different triggers or motivators correspond to differences in the self-help behaviour produced. If so, then what they observe is not really alternative triggers for the same outcomes (multiple causability), but alternative triggers for alternative outcomes. The same might well be so for multiple realisability. It could be that what Wendt observes is not really alternative realisers for the same systemic dispositions, but alternative realisers for alternative dispositions.

The strategy Most and Starr seem to propose is this: when faced with multiple causability (different micro-level causes for a single kind of outcome), we should analyse the outcome with a finer grain of analysis, to determine whether it varies in ways that correspond to the variations in causes. This is essentially what localisation is, and it is perfect for dealing with multiple realisability.[104] If we try to reductively

[103] Most and Starr, 'Foreign Policy Substitutability', p. 402.

[104] See, for example, Jaegwon Kim, 'Multiple Realization and the Metaphysics of Reduction', *Philosophy and Phenomenological Research*, 52:1 (1992), pp. 1–26; and William Bechtel and Jennifer Mundale, 'Multiple Realizability Revisited: Linking Cognitive and Neural States', *Philosophy of Science*, 66:2 (1999), pp. 175–207.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

explain a multiply realised macro-level causal role, we will find multiple mechanisms sufficient to realise that role. But realisers usually differ according to context. Multiple realisation is generally found in dispositional properties that instantiate in different species or kinds of systems.[105] Pain and hunger, mental states that instantiate in everything from humans to octopi, are the most commonly cited examples. Localisation involves: (i) first splitting up a multiply realisable macro-property into subtypes according to each context or species of instantiation, (ii) then linking each subtype to a realiser. This leaves us with a set of context-specific, localised mechanistic explanations, to replace the single macro-level generalisation we started with. If multiple realisation is context-dependent, localisation should be successful.

As an example, consider Wendt's 'Lockean' culture of anarchy, one of the three varieties of collective order he identifies in *Social Theory*, and the one which seems to most resemble the current world system. Wendt argues that the Lockean culture generates behavioural tendencies like balancing behaviour, respect for sovereignty, less concern for relative gains, constrained warfare, and a low death rate amongst states.[106] He concedes that this culture's behaviour- and interaction-shaping role is realised at the micro-level by a common belief amongst states that they are one another's rivals. But no particular arrangement of beliefs is necessary for a Lockean culture or the tendencies mentioned. It does not really matter if some of the states are wrong in their beliefs, or believe something else entirely. It also does not matter exactly what states or people inhabit the system, or even whether there is a full turnover in membership.[107] This culture will still programme for the tendencies with which it is associated. However, if the common belief set underlying this culture ever has varied in a notable, *consequential* sense, then this would have made some observable difference to the mentioned tendencies. For example, the exact distribution and strength of states' rivalry-beliefs has probably changed over time. This, in turn, will have made differences to the exact prevalence of the abovementioned tendencies themselves at different times – like slightly more or less balancing behaviour, war, dissolutions of states, and so on. It may even have meant that certain of these tendencies were not present. Accordingly, we would find that these tendencies – which constitute Lockean culture's causal role – have varied in a way that corresponds with differences in the common belief set underlying this culture. What we are looking at, then, are different realisation-bases for different kinds (or grades) of Lockean culture, not multiple realisation. This is where localisation comes in. Localisation would involve: (i) splitting up (that is, decomposing) the tendency-set into a set of discrete tendency-subsets, along the lines suggested (that is, according to their prevalence or the absence of some), then (ii) localising each to the particular historical period with which it corresponds. Reductive explanation is then a real possibility.

Whether localisation will ultimately work is a matter for further investigation. Nonetheless, I think the above example highlights a basic problem with the multiple realisability argument. Philosophers of multiple realisability typically assume that micro-level properties are differentiated in a more 'fine-grained', variation-sensitive way than macro-level properties.[108] In my view, Wendt and Waltz do the same. But

[105] See Putnam, 'Mental States'; and Kim, 'Multiple Realization'.
[106] Wendt, *Social Theory*, pp. 279–85.
[107] Ibid., pp. 162–4. See also Wendt, 'State as Person', pp. 299–301.
[108] Bechtel and Mundale, 'Multiple Realizability Revisited'.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

when the same, fine grain of analysis is applied at both levels – as in the example above – appearances of multiple realisability tend to wash away, clearing the way for mechanism-based explanation.

Obstacle 3: emergence

Another important obstacle to the microfoundational strategy is what is often referred to as emergence. A number of IR theorists argue that structures are emergent from their agent-level foundations, in the sense of being more than mere 'sums of their parts'.[109] If true, the reductive, agent-based ontology implied in microfoundationalism is fundamentally incomplete. Wendt is at the forefront of this attack as well, and offers several arguments. In his earliest contribution on the matter, he argued that structures are sets of 'internal relations', which constitute properties of their *relata* (states).[110] Later, in *Social Theory*, Wendt argues that structures are institutions of collective knowledge instead, but which also stand in a constitutive relation with agents, and determine certain of their properties.[111] In more recent work, Wendt argues that structures are holographic projections of individual consciousness, which stand in relations of quantum 'entanglement' with agents.[112] All these arguments concern relations of some kind, assert that structures constitute important properties of agents, and suggest that structures do not reduce to their microfoundations.

This line of argument is much more ambitious than that of multiple realisability. The latter is normally coupled with a reductive view of ontology. Most of its philosophical proponents agree with principles like physicalism and supervenience. Their view is that there is only physical stuff in the world, but we cannot explain everything in physical terms because of multiple realisability. The implication is that structures are irreducible in an epistemological, but not ontological, sense. Some refer to this as 'epistemological emergence', to distinguish it from the 'ontological emergence' referred to in arguments like Wendt's.[113]

There are, I believe, a number of problems with Wendt's arguments, but I do not have the space to provide a full critique here. Instead, I want to focus on a particular way of dealing with emergence, in which mechanism-based explanations play a central role. Wendt's arguments rely on a set of relatively abstract examples, in which he tries to demonstrate two main things: (i) that agents depend, for certain of their properties, on the existence of structures; and (ii) that structures, for their part, do not entirely depend on agents for their own existence. His arguments, like many others against ontological individualism in social science, suggest that structures do not wholly supervene, or depend, on agents.[114] There is an epistemological void

[109] See, for example, Jervis, 'System Effects', pp. 12–17; Wight, *Agents, Structures*; Leon, 'Reduction'; and Jonathan Joseph, 'The International as Emergent: Challenging Old and New Orthodoxies in International Relations Theory', in Jonathan Joseph and Colin Wight (eds), *Scientific Realism and International Relations* (New York and Basingstoke: Palgrave Macmillan, 2010), pp. 51–68.

[110] Wendt, 'Agent-Structure Problem'.

[111] Wendt, *Social Theory*. See also Wendt, 'State as Person'.

[112] Alexander Wendt, 'Flatland: Quantum Mind and the International Hologram', in Mathias Albert, Lars-Erik Cederman, and Alexander Wendt (eds), *New Systems Theories of World Politics* (New York and Basingstoke: Palgrave MacMillan, 2010), pp. 279–310.

[113] See Michael Silberstein and John McGeever, 'The Search for Ontological Emergence', *The Philosophical Quarterly*, 49:195 (1999), pp. 182–200.

[114] See, for example, Gregory Currie, 'Individualism and Global Supervenience', *The British Journal for the Philosophy of Science*, 35:4 (1984), pp. 345–58; and Epstein, 'Ontological Individualism Reconsidered'.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

in these arguments, however. Wendt and the other authors mentioned do not tell us how ontological emergence is to be proven. Asserting that structures do not supervene or depend on agents is one thing, proving it in practice is another. What sorts of empirical indicators should we be looking for to prove or disprove emergence? In my view, the indicators need to be causal. To prove that structural properties are more than just collections of micro-level properties, it needs to be shown that they play causal roles distinct from those of their realisation bases. In other words, it needs to be shown that structures' parts are insufficient to realise their causal roles. To disprove supervenience specifically, emergentists would have to show that structural factors play causal roles that can differ even while their realisation bases remain unchanged. I have adapted this criterion from what is known as the 'causal theory of properties' in metaphysics. This theory states that if two properties bestow the same causal powers on things, we should infer that they are really the same property.[115] Applied to the problem at hand, it suggests that we can only infer a structure is more than its microfoundations if it plays a causal role that distinguishes it accordingly. If not, we should infer that they are really one and the same, and there is no ontological emergence. Why? We are only able to perceive and come to know of things and properties by the way they behave and impose themselves on our senses. Thus, as Armstrong argues, if 'things act differently in the same circumstances, then we attribute to them a different nature'.[116] We really have nothing to go by other than causal role, when we are talking about perceptually remote things like social structures. Thus, it would be unreasonable to hold that a structural factor was distinct from its microfoundations if both had the same detectable causal role in all like-circumstances. It would also be unreasonable to hold that a structural factor was nothing more than its microfoundations, if the two played different roles in identical circumstances.

This is essentially inference to the best explanation. Wendt actually uses inference to the best explanation to justify a very similar principle. He argues that we can infer that theories of unobservable objects refer to real things and processes from the fact that they successfully predict observable behaviour.[117] The best explanation for why a given theory predicts observable behaviour is that the theory is true and refers to something real. Wight suggests essentially the same thing when he asserts, quoting Bhaskar, that structures' '*causal power* establishes their reality'.[118]

To establish that structures are emergent from their microfoundations, I believe they need to be shown to satisfy this criterion – that is, shown to play distinct causal-explanatory roles. How can this be established? Certainly not *a priori*, as Wendt, Wight, and others seem to be doing. It can only be established in practice, by the *failure* of mechanism-based explanations. If structures really are ontologically emergent, then we will not be able to explain them in terms of their component mechanisms, and microfoundationalism will fail. Only once this strategy has failed in practice will we be able to confidently dismiss its ontology. As it stands, I do not

[115] For different versions of the theory see David M. Armstrong, *Universals and Scientific Realism, Volume 2: A Theory of Universals* (Cambridge: Cambridge University Press, 1978), pp. 43–7; and Sydney Shoemaker, 'Causality and Properties', in *Identity, Cause, and Mind* (Cambridge: Cambridge University Press, 1984), pp. 206–33.

[116] Armstrong, *Universals*, p. 45.

[117] Wendt, *Social Theory*, pp. 62–5, and 215–7.

[118] Wight, *Agents, Structures*, p. 46.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press

think that systemic research thus far has established that ontological individualism is wrong. Most structural factors with established or hypothesised explanatory roles, like polarity, polarisation, and norms, are already defined in unit-level terms – that is, as collections of unit-level factors. It does not seem necessary that, for any of these factors to play the roles they have been attributed, they need to be anything 'more' than their microfoundations. In any case, these are only things that can be established by attempting mechanism-based explanations. Microfoundationalism may ultimately fail, but in the meantime, there are plenty of good reasons to search for mechanisms.

Conclusion

In this article I have argued that there are good prospects for further exploration and development of the microfoundational strategy in IR theory. Mechanism-based explanations seem to be both feasible and desirable. They appear feasible, because the microfoundational strategy has ways of getting around obstacles like multiple realisability and the role of context. They are desirable, particularly from the point of view of systemic theory, because they give structures and structural theories microfoundations, and could help bring structure back to the centre of discussion in IR theory.

With the obstacles discussed above out of the way, or at least contained, the microfoundational strategy may very well be able to resolve the agent-structure problem. Its success in this regard is a matter for further investigation, however. Other obstacles need to be addressed. What can be determined at this stage though is that even if successful, the strategy would not mean an end to structural research. As we have already seen its solution does not involve giving explanatory priority to lower levels of analysis, or eliminating structural theories. To be sure, it is a reductive strategy, and reductionism is sometimes thought of as an eliminative thesis in philosophy. But microfoundational explanations will not make structural theories of IR redundant. As argued above, they actually have the opposite effect, because they help remove the mystery surrounding structure. From a microfoundational point of view, structural explanations are even more useful than many structural theorists suggest. They are useful because they offer what reductive theories cannot – a more general, parsimonious image of world politics. Mechanism-based explanations, on the other hand, give us depth and detail at the cost of elegance. As Sober has argued, there is 'no objective rule concerning which is better', and any good science should aim for both.[119] Thus, we still need structural research in our explanatory repertoire. This article's point is just that microfoundational research is equally important.

[119] Elliott Sober, 'The Multiple Realizability Argument against Reductionism', *Philosophy of Science*, 66:4 (1999), p. 550. See also Jackson and Pettit, 'Explanatory Ecumenism'.

https://doi.org/10.1017/S0260210513000156 Published online by Cambridge University Press