

## Using Counterfactuals to Investigate Causality in International Relations

Benjamin Mueller, London School of Economics and Political Science (B.D.Mueller@lse.ac.uk)

*I present an argument for the use of counterfactuals as a means of uncovering complex causation. The concept of causal-counterfactuals is introduced and discussed. I proceed by explaining the two analytical benefits of a causal-counterfactual re-opening of the past:*

- 1. To establish why leaders arrived at given choices over others, and to investigate the causal links between said choices and outcomes in the international system;*
- 2. To test the relative influence of agency, structure and contingency on the international system.*

*The mode of counterfactual reasoning presented here partially overlaps with neopositivist understandings of causal analysis, but is shown to provide a superior basis for the study of complexity in the international system. It is suggested that this has positive ramifications for the utility of counterfactual analyses to policymakers. Preliminary findings that this methodological approach has yielded regarding our understanding of the end of the Cold War are briefly presented.*

### *Introduction*

When it comes to the end of the Cold War, the gulf between IR's predictive aspirations and reality is acute. Few in the IR anticipated an end to the conflict when it was already in the process of drawing to a close in the 1980s.<sup>1</sup> As a result, theorists from across the paradigmatic spectrum dove deep into the conceptual arsenal of their respective analytical approaches to account for what had happened, after it happened. Not a small number of scholars earnestly proclaimed the nature of events to have been such that no other outcome than the one that occurred was possible.<sup>2</sup> Declaring the inevitability of an event post hoc, after failing to predict it in the first instance, casts immediate doubt on the worth of the theory that gives rise to such an assertion. It

---

<sup>1</sup> "Not only did almost nobody in politics or academia predict [the end of the Cold War], most forecasts pointed in the opposite direction of what actually happened. And most false predictions followed logically from core assumptions of major international relations theories." in Grunberg, Isabelle and Risse-Kappen, Thomas. 'A Time of Reckoning? Theories of International Relations and the End of the Cold War.' In Allan, Pierre and Kjell Goldmann (eds.) *The End of the Cold War – Evaluating Theories of International Relations*. Dordrecht: Martinus Nijhoff Publishers, 1992. In the same vein, it has been asserted that "The reorientation of Soviet foreign policy under Mikhail Sergeievich Gorbachev and the East-West reconciliation it brought about constitute a formidable challenge to international relations theory. Neither realists, liberals, institutionalists nor peace researchers recognized beforehand the possibility of such momentous change, and they have all been struggling to find explanations consistent with their theories." Lebow, Richard N. and Thomas Risse-Kappen. *International Relations Theory and the End of the Cold War*. New York, NY: Columbia University Press, 1995

<sup>2</sup> See, for example: Brooks, Stephen and William Wohlforth. 'Economic Constraints and the Turn Toward Superpower Cooperation in the 1980s.' In Olav Njølstad (ed.) *From Conflict Escalation to Conflict Transformation: The Cold War in the 1980s*. London: Frank Cass, 2004; Oye, Kenneth A. 'Explaining the End of the Cold War: Morphological and Behavior Adaptations to the Nuclear Peace?' In Lebow, Richard N. and Thomas Risse-Kappen (ed.) *International Relations Theory and the End of the Cold War*. New York, NY: Columbia University Press, 1995; Waltz, Kenneth N. 'The Emerging Structure of International Politics.' *International Security* 1993 Vol. 18

seems the temptation to make predictive statements about world politics is irresistible, even when the prediction concerns the past.<sup>3</sup>

Many scholars of IR study the workings of the international system from a systemic perspective. Different schools of thought emphasise various research themes – realists focus on the consequences of anarchy and the correlation of national interests in the system; liberals on the interaction of domestic processes, institutions and ideas; constructivists on the power of social dynamics, norm-generation, and knowledge communities – and a great deal of time is devoted to empirically pit these research programmes against each other in different contexts. This is a perfectly justifiable way to conduct research: it yields contrasting perspectives on world politics, which offer a coherent set of competing systemic approaches to understand international affairs.

In an effort to understand the workings of the international system better, all these theories place great value on the study of social aggregates. What emerges are big-picture, macro-structural accounts. By contrast, there is little IR research concerning how leaders, through their idiosyncrasies and decisions, affect international politics, not in the wider social context but in the private setting of their decision-making environment. I maintain that investigating the policy dilemmas and uncertainties that leaders faced when they attempted to end the Cold War is key to understanding how and why these political dilemmas were resolved, and the extent to which leaders can drive change.

*Re-opening the past allows for an examination of the links between decisions and outcomes in the international system*

The particular methodological approach I present and explain here is a counterfactual analysis of decision-making: using the menu of choices available to leaders at a given time to study how events could—or could not—have unfolded differently. Through this process I try to re-open the policy alternatives open to leaders in the US and the Soviet Union in 1980s, re-visit what are now facts but were then unresolved problems, focus on the options policymakers faced, trace their decisions and explore how different

---

<sup>3</sup> For example, John Mearsheimer has argued that structural neorealism best explains the absence of a major power war during the Cold War through its modelization of a bipolar, stable world. Consequently, he predicted that following the breakdown of this model and the emergence of a multipolar world, the prospects for inter-state wars would increase, particularly in Europe, where the power vacuum left behind by the implosion of the USSR would bring intra-European security concerns back to the fore and usher in a new period of instability. The perhaps most extraordinary outgrowth of this entirely failed set of predictions was Mearsheimer's policy advice that "the US should encourage the limited and carefully managed proliferation of nuclear weapons in Europe." See Mearsheimer, John J. 'Back to the Future: Instability in Europe after the Cold War.' *International Security* 1990 Vol. 15

choices could have brought about different outcomes.<sup>4</sup> As Hugh Trevor-Roper has put it, by studying the political dilemmas of yesterday, by restoring the past's 'lost uncertainties, to reopen, if only for an instant, the doors which the *fait accompli* has closed,' we can draw useful lessons from history.<sup>5</sup> It is all too easy to see a pattern in the myriad of actual and potential events of yesteryear, and draw a line through the dots that just so happens to take us straight to the present. But that is ultimately an instance of hindsight bias. Again, Trevor-Roper formulated it elegantly: 'In retrospect, we read the signs, select the evidence, and complacently predict what has already only too visibly happened.' This predisposes the researcher to analytical certitude and complacency. The study of uncertainty – while fraught with cognitive biases of its own – may be more fruitful than the post hoc rationalisation of the unpredictable. As put by Louis Menand:

“If you look backward, the dots that lead up to Hitler or the fall of the Soviet Union or the attacks on September 11th all connect. If you look forward, it's just a random scatter of dots, many potential chains of causation leading to many possible outcomes. We have no idea today how tomorrow's invasion of a foreign land is going to go; after the invasion, we can actually persuade ourselves that we knew all along.”<sup>6</sup>

To reconstruct decision-making and 'imagine' different choices and outcomes, serves not just to make us aware of the underlying indeterminacy of world history (at least on the basis of our current epistemological tools). Counterfactuals can also be used to shed light on causality, a subject that IR handles rather reluctantly. Stephen Walt has noted that the rise of formal theory in political science and IR has been coloured by a strong leaning toward a probabilistic-mathematical approach to causation. This methodological approach proceeds along the following lines: a model is mathematically formulated, dependent and independent variables are specified, data pertaining to each variable is obtained, and the correlation examined.<sup>7</sup> The aim of such research is ultimately to a) test the strength of the correlation between two variables and b) establish whether the two variables co-vary in the manner predicted by the researcher. The trouble is that correlation need not mean causation. No matter how mathematically sophisticated a model may be, no matter how many datasets

---

<sup>4</sup> See, for instance, Mueller, Benjamin. 'A Counterfactual Estimate of the Impact of Leadership and Diplomacy on the End of the Cold War.' Paper to be presented at ISA, 29 March 2014, Toronto

<sup>5</sup> Trevor-Roper, Hugh. 'History and Imagination.' In Lloyd-Jones, Hugh (ed.). *History & Imagination*. London: Holmes & Meier Pub, 1982

<sup>6</sup> Menand, Louis. 'Everybody's an Expert – Putting Predictions to the Test.' *The New Yorker*. December 5 2005. Web. Accessed on 23 March 2014 at <[http://www.newyorker.com/archive/2005/12/05/051205crbo\\_books1](http://www.newyorker.com/archive/2005/12/05/051205crbo_books1)>

<sup>7</sup> Walt, Stephen M. 'Rigor or Rigor Mortis? Rational Choice and Security Studies.' *International Security* 1999 Vol. 23

indicate that the variables do behave in the way suggested by the model: probabilistic empirical testing leaves the actual causal logic of the model untouched – it can only find evidence for or against it.

*Thinking about causation counterfactually opens up unconventional perspectives on causal analysis in IR*

There is a school of philosophy that analyses causality in terms of counterfactual conditional statements. All causal thinking, it is asserted, is counterfactual thinking.<sup>8</sup> The general structure of the argument runs as follows: if A and only A is said to cause B, it follows that in the absence of A, there is no B. Max Weber used this reasoning to call for the counterfactual investigation of causal processes in history:

The judgment that if a single historical fact in a complex of historical conditions [had been] missing or altered this would have brought about a ... divergent course of historical events [is crucial] in the determination of the 'historical significance' of that fact. The first – and crucial – [abstraction involved in causal analysis] is just this: that among the actual causal components of a course of events, we think of one or several as altered in a certain direction, and we ask ourselves whether, under the changed conditions ... the same – or what other – outcome was “to be expected.”<sup>9</sup>

In other words, in order to evaluate a claim about the degree to which a given factor 'favoured' a given historical outcome, the hypothetical substitution of either the outcome or the suggested cause will yield insights into whether the causal connection holds.<sup>10</sup> Such a counterfactual investigation of causes can explain why a particular historical outcome came about instead of another. The combination of causal and counterfactual reasoning yields what Ringer terms “a dynamic vision of alternate paths of historical change.”<sup>11</sup>

This mode of analysis raises the question of what exactly is meant by a cause.<sup>12</sup> Definitions of causation can be grouped, broadly speaking, into causes as: 'probability-raisers', 'events linked to other events through effect mechanisms' or 'necessary

---

<sup>8</sup> See Williamson, Timothy. 'Philosophical Knowledge and Knowledge of Counterfactuals.' In Beyer, Christian and Alex Burri (eds.) *Philosophical Knowledge – Its Possibility and Scope*, Amsterdam: Rodopi, 2007; Lewis, David. 'Causation'. *Journal of Philosophy* (1973): pp. 556-67; Collins, John, Hall, Ned and Paul A. Lewis. *Causation and Counterfactuals*. Cambridge, MA: MIT Press, 2004

<sup>9</sup> Weber, Max. 'Kritische Studien.' pp. 274-275. Cited in Ringer, Fritz. 'Max Weber on Causal Analysis, Interpretation, and Comparison.' *History and Theory* 2002 Vol. 41

<sup>10</sup> Ringer, p.162

<sup>11</sup> Ibid, p. 167

<sup>12</sup> For a thorough discussion, see Kurki, Milja. 'Causes of a divided discipline: rethinking the concept of cause in International Relations theory.' *Review of International Studies* 2006 Vol. 32, pp. 189-216; also Suganami, Hidemi. 'Philosophy as midwife of science? The role of philosophical analysis in causal investigations in the study of world politics.' Draft paper.

conditions for effects.<sup>13</sup> The first, probabilistic view of causation as observed regularity is informed by Hume’s position that causal processes can be inferred through experimental covariation. This forms the basis of neopositivist science’s concern with the systematic replication of purported causes, so as to measure their effects.<sup>14</sup> Whatever definition one may hold of causation, Grynaviski has shown that a common denominator in causal analysis is the type of question asked, namely so-called ‘why-questions.’ King, Keohane and Verba “argue that research questions should request explanations (e.g. causal responses) and Alexander Wendt explains that these requests for explanation take the form of ‘why-questions.’”<sup>15</sup> ‘Why *x*’ questions, however, are open-ended. As David Lewis has maintained, by asking what caused *x*, we are in effect requesting the entire list of causal events that took place in the run-up to *x* – where *x* “is the culmination of countless distinct, converging causal chains.<sup>16</sup> That isn’t very useful when it comes to figuring out which causes were significant. The way to narrow down requests for causal explanation, as Grynaviski suggests, is by asking *contrastive* why questions: ‘Why *x* rather than *y*?’ This type of question allows researchers to zoom in on a specific causal riddle, thereby elevating significant causes over other, trivial ones: “Contrastive why-questions highlight variation and puzzling phenomena and enable detection of salient causes.”<sup>17</sup> Asking ‘Why *x* rather than *y*’ amounts to a counterfactual statement: implicit in the inquiry is a look at the causal logic that underpins *x* vis-a-vis the causal process that would have brought about *y*.

Linking up a chain of contrastive-why questions provides “the best approach to understanding the causal dynamics of historical events in their richest details.”<sup>18</sup> In the words of Jackson, under this methodology, “researchers trace and map how particular configurations of ideal-typified factors come together to generate historically specific outcomes in particular cases.”<sup>19</sup> Such an approach demands the in-depth historical study of the interactions between political elites on both sides during that time. The method advocated here is the construction of a counterfactual-causal chain of decision-making related events at the end of the Cold War, with the aim of narrating the causal dynamics at play in this recent instance of system change in IR. This should yield relevant cause-effect relationships and bring them into a timed sequence, instead of abstracting causes and breaking them up into law-like

---

<sup>13</sup> Grynaviski, Eric. ‘Contrasts, counterfactuals, and causes.’ *European Journal of International Relations* 2013 Vol. 19

<sup>14</sup> Jackson, Patrick T. *The Conduct of Inquiry in International Relations: Philosophy of Science and its Implications for the Study of World Politics*. New York, NY: Routledge, 2010, p. 69

<sup>15</sup> Grynaviski, p. 7

<sup>16</sup> Lewis, David. ‘Causal explanation.’ In *Philosophical Papers, Vol. II*. New York: Oxford University Press, p. 214

<sup>17</sup> Grynaviski, p. 10

<sup>18</sup> *Ibid*, p. 12

<sup>19</sup> Jackson, p. 114

generalisations.<sup>20</sup> The methodological tool is the “disciplined use of counterfactuals – not as a way of elucidating the implications of law-like generalizations, but as a way of imagining alternative historical trajectories that might have led to different outcomes than that actually observed.”<sup>21</sup>

Causal-counterfactual narratives can make the nature of change in IR – its contingencies, idiosyncrasies, unpredictability – meaningful. The idea is to make sense of the vast petri dish of world politics by causally tying together the various events that in aggregation produce ‘history’. This is not in opposition to established research methods: it simply produces a different focus, one which complements our increasingly rich portfolio of structural explanations for world politics.

### *Fleshing out a counterfactual-causal approach to the study of international relations*

Narrative history has its own problems. Chronicling events reveals nothing about the patterns and regularities that can be observed across time. Unlike historians, the IR scholar should still seek out what Jon Elster has termed “plausible, frequently observed ways in which things happen.”<sup>22</sup> There is a danger that in the process of historical research, the scholar is over-awed by the mosaic texture of world history and thereby loses sight of the need to fit his work into broader causal structures. A historical approach must not ignore structural theorising. Individuals in positions of political power are constrained in their dealings by the structural environment within which they operate. At the very least, this sets out a sometimes clear and sometimes less clear material framework, which in turn informs the menu of choices individuals are faced with. Gorbachev, on assuming power, realised or already knew that the Soviet Union would struggle to afford entering a renewed arms race and meet Reagan’s Strategic Defence Initiative head-on, without having to pay an enormous – in all likelihood unpalatable – opportunity cost. Clearly, this structural reality was a key factor influencing his increasingly fundamental conviction that the only way to end the arms race was through arms treaties that substantively reduced the USSR’s military burden.

The crucial question, however, is precisely *how* and *why* the USA and the Soviet Union signed the treaties they did: what were the negotiation processes that finally produced success rather than degenerating into the mutual recriminations that had stalled arms control efforts since the collapse of SALT II in 1979? The details of the

---

<sup>20</sup> For more on neopositivism’s inability to do ‘time’, see Hom, Andrew H. ‘World’s enough, and less time: narrative reasoning about a time-bound world.’ (forthcoming)

<sup>21</sup> Jackson, p. 115

<sup>22</sup> Elster, John. *Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press, 1989

interactions that produced the known outcome allow for a counterfactual-causal exploration of what happened and why it did, relative to what other options were on the table. To argue that the material reality was such that Gorbachev was left with little choice – a favoured realist argument – radically simplifies this momentous development. These kinds of conceptual shortcuts leave realist analysis bereft of the depth that is needed to understand how a six-decade long superpower confrontation – with more than its fair share of close nuclear shaves – ended within a few years, without a single shot being fired between the two adversaries. Naturally, Gorbachev was under pressure to act on and improve the USSR's situation. Equally clear, however, is the fact that the policy paths he chose were not foreordained: Gorbachev faced real choices and real alternatives. Why did he not pursue perestroika without glasnost, for example? It is imperative to trace out how Gorbachev and Reagan arrived at their respective policy choices that brought the Cold War to a peaceful conclusion.

After events have happened, it is relatively easy to erect a narrative suiting a particular theory that makes decisions look preordained. In actual fact, the uncertainty leaders face when they deliberate on a decision is, more often than not, vast and preceded by intense policy debates. The 'right' answer (or – in enough cases – the 'wrong' answer) may appear obvious in retrospect, but focusing solely on the chosen policy path fails to take into account the reality of decision-making at the time. When a system is in crisis, actors need to respond. The solution to the crisis will be far from clear and various options will be forward, which may seem equally adequate at the time given the lack of complete knowledge. The formulation of a policy response by leaders is as relevant as the actual decision itself if we wish to understand outcomes in IR. Almond and Genco have explained this understanding political reality well, arguing it “consists of ideas – human decisions, goals, purposes – in constant and intense interaction with other ideas, human behaviour and the physical world. At the centre of this complex system are *choices* and *decisions* ... that is the heart of politics.” [Emphasis added]<sup>23</sup>

Karl Popper, the father of critical rationalism, realised that his main argument concerning scientific methodology – that science consists of the systematic testing of hypotheses through experimentation – could not be applied to the study of the past, since social events and trends are caused by initial conditions and are, as such, non-linear. And yet, Popper insisted, there can be causal explanation in history without the need to rely on deductive methods: “There are countless possible [initial] conditions; and in order to be able to examine these possibilities in our search for the

---

<sup>23</sup> Almond, Gabriel A. and Stephen J. Genco. 'Clouds, Clocks, and the Study of Politics.' *World Politics* 1977 Vol. 29

true conditions of a trend, we have all the time to try to imagine the conditions under which the trend in question would disappear.”<sup>24</sup> Natural scientists can design deductive experiments to test their postulated causal laws, historians can reach conclusions about conditions and antecedents through inductive means such as counterfactual reasoning.

### *Counterfactual parameter #1: Plausibility*

As seen, counterfactual thinking has a rich philosophical pedigree. Isaiah Berlin, too, called on scholars to establish the possible courses of action open to human beings in the present *and* the past. He called for “the placing of what occurred (or might occur) in the context of what could have happened (or could happen), and in the demarcation of this from what could not.” This issue of the demarcation between the actual, the possible and the impossible is crucial to relevance of a counterfactual. How do we prevent counterfactual analysis from degenerating into fanciful flights of intellectual self-indulgence, where wild ‘what if’-dreams are pursued that are so far removed from the actual course of events that their absurdity precludes any possibility of ever contributing to anything other than a lively after-dinner conversation? The key, Berlin recognised, is the plausibility of the counterfactuals we construct:

“When an historian, in attempting to decide what occurred and why, rejects all the infinity of logically open possibilities, the vast majority of which are obviously absurd, and, like a detective, investigates only those possibilities which have at least some initial plausibility, it is the sense of what is plausible – what men, being men, could have done or been – that constitutes the sense of coherence with the patterns of life.”<sup>25</sup>

To avoid the danger of slipping into an infinite regress and construing all available paths as possible it is important to limit the enquiry to plausible scenarios. The distinction between what did happen and what could *plausibly* have happened is crucial. Ferguson has applied this standard to counterfactual research: plausibility is ensured by considering as probable only those alternatives which can, on the basis of contemporary evidence, be shown to have actually been considered by the relevant actors at the time. As Ferguson has put it: “What we call the past was once the future; and the people of the past no more knew what their future would be than we know our own. All they could do was consider the likely future, the plausible outcome.”<sup>26</sup> This

---

<sup>24</sup> Popper, Karl R. *The Poverty of Historicism* London: Routledge & Kegan Paul, 2002, p. 119

<sup>25</sup> Berlin, Isaiah. *The Proper Study of Mankind: An Anthology of Essays*. London: Chatto & Windus, 1997, p. 56.

<sup>26</sup> Ferguson, Niall. *Virtual History: Alternatives and Counterfactuals*. London: Macmillan, 1997, p. 82.



plausibility safeguard, however, applies only to counterfactuals that deal with potential alternative decisions taken by leaders.

Whilst alternative policy choices are worthy of study for the reasons outlined above, another use of counterfactuals is determining the relative causal influence of agency, structure and contingency: that is, by counterfactually replacing actors, or assuming alternative structural conditions, or changing timings, arguments can be constructed that show the causal influence of and interrelationships between each class of variable on particular situations. Such counterfactuals fall outside the remit of Ferguson's plausibility safeguard and require another methodological parameter to become operational.

### *Counterfactual parameter #2: Scope*

The scope of the statement under investigation defines the depth and breadth of the scenario needed to construct the counterfactual, and by extension its plausibility. John Mueller has crafted an interesting counterfactual proposition regarding the influence of nuclear weapons on the war-proneness of the international system. Using the rational actor model as a covering law to account for leaders' decisions about going to or refraining from war, he traces out the development of post-WW2 history, supposing the absence of nuclear weapons.<sup>27</sup> He argues that what kept the peace between the US and the USSR was not the spectre of nuclear warfare, but the fact that the cost of going to war had already increased to such an extent – what with the horrors of World Wars One and Two – that nuclear weapons were not a causal difference-maker during the unfolding of political history since 1945. However, counterfactual reasoning across such a long-time period – changing one element of world history and then re-writing decades of history on the basis of that change only, but keeping all other developments as they had been – is difficult. As Tetlock and Belkin put it, “it is not at all clear that cotenability obtains between the counterfactual antecedent of a non-nuclear world and any connecting principle that posits the occurrence of the Cuban missile crisis in 1962.”<sup>28</sup> In other words, one cannot simply counterfactually change one fundamental fact of world politics and then ignore the ripple effects of said change. The more substantial the change posited by the counterfactual, and the longer the period over which history is then re-played, the more complex the process of spelling out one's alternative becomes, and the more amenable the counterfactual becomes to the criticism that it is a work of fiction rather

---

<sup>27</sup> Mueller, John. ‘The Essential Irrelevance of Nuclear Weapons: Stability in the Postwar World.’ *International Security* 1988 Vol. 13

<sup>28</sup> Tetlock, Philip E. and Aaron Belkin. *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives*. Princeton, NJ: Princeton University Press, 1996, p. 23

than a piece of plausible analysis. Williams, a philosopher who has studied the relationship of counterfactual reasoning to metaphysical modality and knowledge, maintains: “Counterfactuals whose antecedents involve small departures from the actual world are easier to evaluate.”<sup>29</sup>

*Applying counterfactual-causal thinking to the end of the Cold War*

Because of the speculative reasoning involved, it seems that key measures to maintain the plausibility of a counterfactual are a) keeping the re-write of history to a minimum, b) basing the unfolding of subsequent events on some kind of evidence and c) if ‘removing’ an element of history, be it an actor, an invention, or an event, all ensuing analysis needs to adequately take the inter-connectedness of causes into account, to pre-empt accusations of flippancy. That, of course, is easier said than done.

For example, when it is asserted that ‘the relative decline in power of the Soviet Union caused the end of the Cold War’, imagining the opposite – that Soviet power did not decline – involves a huge counterfactual transformation of actuality: a whole new world would have to be imagined in which Soviet communism did not suffer from prolonged economic weakness from the 1970s onwards. The greater the reach of the imagined scenario, the more controversial the supposition becomes. By contrast, narrowly-defined and well-specified counterfactuals are far more useful in assessing the worth of causal statements. Take, for example, the following assertion: ‘The Strategic Defence Initiative proved a crucial escalation in the arms race which forced Gorbachev to concede that the USSR’s ambition to keep up militarily with the United States was unsustainable and led to Soviet arms reductions.’ This causal statement is bounded and well-defined: SDI caused Soviet arms reductions. A counterfactual look at the statement’s strength would entail supposing a Soviet response that differed from the actual response (which was retrenchment): what if the Soviets had reacted with their own version of SDI? More specifically, what would it have taken for this to take place? To construct such a counterfactual, a further set of questions need to be answered: Did the Soviets lack the technological and/or financial resources required? Did Gorbachev idiosyncratically reject a Soviet response in kind? In order to do conduct this analysis, the policy debates within the Soviet leadership surrounding SDI need to be scrutinized: was it Gorbachev who resisted a Soviet SDI, or did others in the Soviet leadership concede the same? Answering these questions reveals whether SDI altered the East-West strategic gambit in favour of the USA because it forced a Soviet hand which it didn’t have, or because Gorbachev’s unique response to this new escalation in the East-West arms race was to withdraw from it.

---

<sup>29</sup> Williams, p. 31

As already noted, the very conceptual underpinning of counterfactuals – the re-imagination of history, a supposition of that which never was – is also the key reason they elicit scepticism among certain parts of the academic community. Even though it serves a clear analytical purpose, the fundamental problem of induction remains: the analyst simply doesn't know what would really have happened in his alternative world. Qua Ferguson, by re-imagining solely decisions, and considering as antecedents exclusively those policy options that participants themselves considered and almost accepted, we can strengthen the plausibility of our counterfactuals, at the price of limiting the enquiry using an arbitrary benchmark: only that which actors have committed to the written record can be considered for counterfactual analysis. This incurs a cost in terms of the completeness of the analysis, but the added benefits of rigour and credibility more than make up for this. The problem would only become insurmountable if there is systematic co-variation between policy debates of which there is no record, and historical outcomes. There is no reason to suspect that this is the case.

*Counterfactuals as a bridge between neopositivism and singular causal analysis*

At this stage, a comment on Jackson's characterisation of counterfactual reasoning within the context of singular causal analysis is in order. According to Jackson, the Weberian approach to counterfactual reasoning stands in fundamental contrast to neopositivist counterfactuals: the latter treat counterfactuals as an experiment of sorts – a thought-experiment – subsumed under the wider epistemological project of falsifying a hypothesis. Jackson states that for neopositivists, causal claims emanating from counterfactuals ought to be systematically arranged so as to generate a set of law-like generalisations, in order to capture causality in the actual world. Weberian causal analysis, by contrast, aims “to identify adequate and coincidental causes by referring to the complex particularity of the case to various ideal-typical statements of analytically general relations,” rather than fleshing out “the logic of a hypothesized relationship, nor the estimation of the various ways that a set of variables taking on discrete values might condition one another in different experimental or quasi-experimental trials.”<sup>30</sup> Jackson goes on to distinguish between adequate, coincidental and incidental causation, arguing that it is the strength of a single causal factor on this scale that is established by singular causal analysis.

It seems, however, that it is not necessary to treat the distinction between neopositivist and singular causal counterfactuals as categorical. There are differences,

---

<sup>30</sup> Jackson, p. 149

naturally. As implied by the name, singular causal analysis organises scholarly thoughts about a particular case; systematic generalisation through the extrapolation of causal findings is not the aim of this process. Because singular causal analysis is ideal-typical in nature, it does not seek to serve a direct contribution to empirical generalisations. This is a result of the mind-world monism espoused by singular causal analysis. Neopositivists' commitment to mind-world dualism – i.e. an ontological belief in the conceptual distinction between the observer and the observed – makes it amenable to systematic, cross-case comparison. Analyticists, by contrast, use ideal-types to study specific cases and can use findings to organise empirical observations into systemic facts, rather than testing their causal chain systematically against empirical data. This is because an analyticist believes the logical independence of a particular causal claim precludes it from being subsumed under general law-like statements. For a neopositivist, by contrast, using multiple case regressions to quantitatively capture the independent impact of a particular variable on a given outcome makes sense: this can establish whether there is consistency between the observer's postulations and the observed world. This method serves no purpose, however, when the postulated causal relationship is immanent (as is the case for the analyticist) rather than exogenous: all that the researcher strives for in the immanent scenario is a logical, case-specific explanation.

Crucially, as Jackson establishes, neopositivism and analyticism share a commitment to phenomenalism.<sup>31</sup> Moreover, they both seek to uncover causal relationships. But while neopositivists aim to falsify hypotheses and generalise law-like relationships in their endeavour to arrive at transcendental, objective truth, analyticists construct narratives, “to organise the empirical material of specific cases into a coherent story that differentiates between analytically general and case-specific factors.”<sup>32</sup> Is this difference really fundamental? The two approaches solve the mind-world problem differently, but are nevertheless both means to uncover causation. Their different positions on the relationship between research and the researched world manifests itself methodologically, whilst the actual ‘analytical stuff’ of both paradigms is causal. Hence both lend themselves to counterfactual analyses – neopositivism with a view to testing theories, analyticism in order to establish convincing causal narratives for specific cases (e.g. was a given decision a function of a leader's personality, or the untenability of other alternatives?). In neopositivist hypotheses, as Fearon has shown, implicit counterfactuals are needed to rule out that a causal link exists between the dependent variables and the error term. An

---

<sup>31</sup> I.e. the doctrine that human knowledge is founded on the realities or appearances presented to the senses.

<sup>32</sup> Jackson, 154

analyticist uses counterfactuals to delineate the causal chain that produced a particular outcome, highlighting paths not taken – and, along the way, explaining why. Similar aims, similar methods, differentiated solely by the two methodologies' respective take on the mind-world problem. Hence a counterfactual analysis like Mueller's combines both elements of neopositivist hypothesis-testing and singular causal analysis' process-tracing. The border between neopositivism and analyticism is porous. For example, the rational actor model underpins a huge swathe of neopositivistic research in economics – despite being an ideal-typical counterfactual model ('If man did not behave rationally, the world would look very different'), not an empirical proposition per se.

The two approaches, both believing in phenomenalism, are not fundamentally different, then. Analyticist methodology does not seek to undermine other modes of analysis but to complement them. Its analyses still rest firmly on the combination of logic and evidence: this allows scrutiny of whether the claims made are plausible on a) logical grounds and b) supported empirically.<sup>33</sup> The methodology remains a social-scientific rather than a reflexivist endeavour, despite not setting out to test one specified hypothesis and delineating the variables at play.

*Causal-counterfactuals provide a more suitable mode of analysing complexity than alternative methods*

The use of singular causal analysis to delineate the unfolding of particular historical trajectories overcomes a key flaw of neopositivism: reductionist-linear models of world politics do a poor job of explaining complex change. A reductionist approach struggles to accurately capture the intricate dynamics that characterise change in international politics. If a system's variables cannot be effectively isolated from each other or from their context, then "linearization is not possible, because dynamic interaction is one of the system's defining characteristics."<sup>34</sup> Nonlinearity has become of great interest to the social sciences of late, accounting for the unpredictability of economic-political developments by incorporating phenomena such as "interconnectedness and context, interaction, chance, complexity, indistinct boundaries, feedback effects,"<sup>35</sup> as opposed to modelling social developments in the mould of Newtonian physics. As already noted, Karl Popper who recognised the importance of initial conditions on the evolution of complex non-linear systems – or

---

<sup>33</sup> De Mesquita, Bruce B. 'The methodical study of politics', Shapiro, Ian, Roger M. Smith and Tarek. E. Masoud (eds.). *Problems and Methods in the Study of Politics*. Cambridge: Cambridge University Press, 2004, pp. 227–47

<sup>34</sup> Beyerchen, Alan. 'Clausewitz, Nonlinearity, and the Unpredictability of War.' *International Security* 1992 Vol. 17

<sup>35</sup> Ibid, p. 82

'trends', as he called them – in his 1957 book, 'The Poverty of Historicism'. Small, contingent events can have infinite consequences in chaotic, non-linear systems. Popper was acutely sensitive to the need to peer beyond actuality and examine the conditions that would have given rise to an alternative actuality ("to imagine a change in the conditions of change",<sup>36</sup> in his words) if we are to gain a genuine understanding of causality. A concern with the non-linearity of change in IR thus entails the need for a research approach that seeks to understand the interactive whole of dynamic system change.

Somewhat surprisingly, perhaps, an early advocate of this kind of holistic causal theory was Clausewitz. According to Beyerchen, Clausewitz deemed it imperative to face "up to the intrinsic presence of chance, complexity, and ambiguity in war [...] For Clausewitz, this is preferable to the risk of being blind-sided by the strictures of a theory artificially imposed on the messiness of reality in the name of clarity [...] His [Clausewitz's] concerns, like those of many scientists wrestling with nonlinear phenomena today, are open systems which cannot be isolated from their environments even in theory, which are characterized by numerous levels of feedback effects, and which need to be grasped realistically as an interactive whole."<sup>37</sup> This is a nice depiction of the ontological basis outlined in this paper.

The end of the Cold War is an episode of systemic change, characterised by a multiplicity of aggregated causal factors. A fruitful way of connecting all these dots together is by telling a causal story that ties up the evolutionary dynamic of the international system in the 1980s.<sup>38</sup> Naturally, a complete replication of the decade cannot be given; the purpose of analysis, after all, is to synthesize events. Given the methodological benchmark applied here to counterfactuals – that of relying on policy alternatives considered but not selected by leaders – a focus on critical junctures of decision-making forms the most straightforward framework within which to construct a causal narrative of the Cold War's end. The outcomes of an analyticist examination of causal dynamics, relying on counterfactual reasoning to establish reasons for the reality of given occurrences over other alternatives, is made tangible by stringing together the cause-effect arrangements deemed to holistically depict the trajectory of events.

*Counterfactual analysis has the potential to generate findings that are of relevance to policy-makers*

---

<sup>36</sup> Popper, Karl. M. *The Poverty of Historicism*. London: Routledge, 1957, p. 124

<sup>37</sup> Beyerchen, pp. 75, 82

<sup>38</sup> For a useful description of technical cause-effect accounts of complex evolutionary processes using narrative methods, see Tilly, Charles. *Why?* Princeton, NJ: Princeton University Press, 2006

Counterfactuals underpin all explanation-seeking analytical thinking. As Hawthorne noted, “an explanation suggests alternatives. [...] the force of an explanation turns on the counterfactual which it implies.”<sup>39</sup> Thus, whether we like it or not, counterfactuals are deeply engrained within the analytical fabric of IR, by virtue of the discipline’s aspiration to explain events in the domain of international relations. Moreover, any kind of goal-oriented decision-making rests on a counterfactual supposition of the cause-effect consequences of one’s choice. Counterfactuals have thus always been a part of IR’s analyses, since they pose a basic causal-analytical device essential to any kind of logical investigation, though this was not made explicit until recently.<sup>40</sup> It is important for IR to begin ‘handling’ counterfactuals in an explicit and transparent manner, so as to harness their analytical value-add and advance IR’s research programme. By attempting to do this in the context of decision-making during the end of the Cold War, my research aspires to pave the way for a clearer understanding of the dynamics of change that drive international relations forward. This could profoundly improve IR’s ability to aid practitioners: in an increasingly complex world, decision-makers need to rely on historically informed judgment to steer their nations successfully through what is a nonlinear-multicausal rather than an linear-axiomatic world.

Jervis has previously noted the suitability of counterfactual analysis for the study of complex systems.<sup>41</sup> Such a system has emergent properties: its units are interconnected such that the system’s characteristics and behaviour cannot be inferred from the individual behaviour of the units within. Changes in one unit or the relationship between any two have ramifications in other units or relationships. In such a system – which presents a realistic ontological grounding for world politics – causation operates in ways that elude standard means of scientific analysis. Counterfactual thinking facilitate the development of our understanding of how elements in complex systems are connected and results arise: well-structured counterfactual scenarios alert us to the possible presence of causal pathways that we could otherwise ignore.

---

<sup>39</sup> Hawthorne, Geoffrey. *Plausible Worlds: Possibility and Understanding in History and the Social Sciences*. Cambridge: Cambridge University Press, 1991, pp. 13-14

<sup>40</sup> Take this famous quote from Thucydides, the first thinker of IR to be recognized as such: “The growth of the power of Athens, and the alarm which this inspired in Lacedaemon, made war inevitable.” This unequivocal causal claim (seemingly in the mould of the ‘necessity’ school of causal thought) contains within it a powerful but implicit counterfactual: absent Athens’ rise, the Peloponnesian war would not have occurred. Athens’ rise was a necessary and sufficient condition for the war: the analysis therefore presumes that agency played no part in causing the war (other than indirectly, through the agency of Athenians bringing about the rise of their city-state) and leaders could not have prevented the war in any event. See Thucydides, *The History of the Peloponnesian War*, Book I, 1.23

<sup>41</sup> Jervis, Robert E. ‘Counterfactuals, Causation and Complexity.’ In Belkin, Aaron and Philip E. Tetlock, *Counterfactual Thought Experiments in World Politics*

By capitalising on counterfactuals' intrinsic link to causal reasoning, and showing how counterfactuals can aid our understanding of events in IR, decision-makers could end up better-placed to make the right calls at world politics' critical junctures of the future. Sound leadership requires an awareness of causal tendencies in the international system, knowledge of the scope and limits of agency, and an accurate, self-reflexive use of counterfactual modality to weigh various scenarios against each other and gain an understanding of the contingencies inherent in each.

### *Placing counterfactual-causal analysis alongside the canon of IR scholarship*

It would be hubristic to call IR theories' explanations of the end of the Cold War 'wrong:' all accounts produce some insights that go some way to clarify what went on in the international system in the 1980s. Realists are right to point out that objective material pressures matter greatly in determining a state's position in the international system. But there is no singular causal direction from material structural developments to changes in policy, and certainly no single policy path open to the decision-maker responding to material pressures. Thus, realism's account remains underspecified. Liberals offer a convincing perspective on how the ideological basis of Communism was eroded by the West's relative success and freedom. But they fail to explain why the Soviet desire for change took on a liberal mantle, when technocratic economic reform along Chinese lines could also have been attempted, along with the maintenance of a repressive state apparatus. Constructivist theories do well to highlight the interaction of agents and structures when identities are re-constituted, and shed light on the process by which norms evolve and permeate the policy-making establishment. However, they do poorly on the causal front, helping to answer 'how possible' questions rather than the 'why questions' that explain what brought about a given outcome over another. Lastly, the cognitive approaches that integrate psychological accounts of how agents adopt new norms don't delimit idiosyncrasy against the external structures that exert unified pressure on all foreign policy agents (e.g., national security imperatives).

### *Summary of argument*

International Relations is a data-poor but fact-rich discipline. In this context, it is useful to argue causal claims counterfactually. To establish a causal link between two events, it needs to be shown that removing the antecedent (the purported cause) changes the consequent (the outcome under investigation). If multiple antecedents bring about the same consequent, said event is redundant as opposed to contingent. If antecedent X can lead to many outcomes {Y, Z, ...}, outcomes {Y, Z, ...} need to be

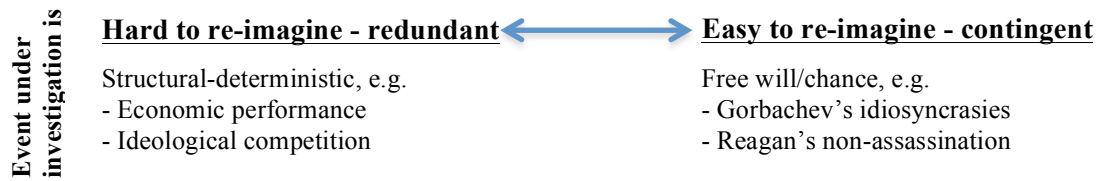


further causally reduced to other antecedents. Where this is not possible, the process under investigation is subject to randomness.

There are two important new contributions my research makes to IR theory and to our understanding of the end of the Cold War.

- 1) Conceptual-substantive: I look at events in the 1980s and build a fact-driven causal narrative. The list of causes that influence the evolution of political events run along a spectrum with two poles (see Fig. 1): located at one end are structural variables with deterministic properties, and on the other end are agents possessing free will as well as chance. Counterfactually probing how and whether events could have unfolded differently allows me to determine how contingent events were: events that are 'hard' to re-imagine are located nearer to the structural-deterministic pole (e.g. hardening of the US stance vis-à-vis the Soviets by the beginning of the 1980). Events that can easily be re-imagined (e.g. a Soviet invasion of Poland in 1980) are contingent and located near to the free will-chance pole (in the case of Poland, Brezhnev's idiosyncrasy was causally significant). When an event turns out to have been contingent, I can explore its wider causal impact by conjecturing what its non-occurrence would have entailed (e.g. no Chernobyl)
- 2) Conceptual-methodological: By deploying counterfactuals to distinguish incidental from coincidental causes I have a tool to assess the relative value of existing theories about the end of the Cold War (broadly divided into structural, domestic-political, ideas-based and leadership theories). Moreover, I can explore the implications of timing and of path-dependencies on the comparative worth of theories (e.g.: the idiosyncrasies of Reagan and Gorbachev causally dominated nuclear disarmament during and following the Reykjavik summit, while structural-realist factors lessened in importance at such highly charged, personal meetings) and explore interaction effects between the theories (e.g. how could Gorbachev [a leader] have turned around the Soviet economy [a structure]?)

**Fig. 1 – Spectrum of causal in/determinacy revealed by counterfactual analysis**



Broadly speaking, my underlying theoretical framework is one that argues against 'big picture' theories and in favour of the fox-like approach advocated by Berlin<sup>42</sup>: my methodology allows picking and choose the theories that best account for given developments. This is done by testing the causal implications behind theories, e.g. by looking at alternative decisions contemplated by leaders but not taken. The narrative approach that underpins this also entails a re-examination of Cold War history, which – given the analytical backdrop of the deterministic-free will/contingency spectrum – yields certain surprising, hitherto underappreciated insights into the operation of a complex international system. Examples of this include the link between the (and highly contingent) Cuban Brigade Affair in the summer of 1979 and the Soviet invasion of Afghanistan, or the unexpected policy continuities between Carter and Reagan as well as the causally relevant differences between the two.

In sum: I use counterfactuals to study the root causes of the end of the Cold War. My methodology essentially amounts to 'predicting the past': studying how events have unfolded versus how they could have unfolded, so as to better differentiate between essential and incidental causes. Counterfactuals suggest causal pathways. The credibility of a postulated cause rests on how convincing its counterfactual inverse is. History is non-repeatable, so reverse-engineering it is enormously challenging.<sup>43</sup> But by bringing about a better understanding of the general properties of the phenomenon under study (how structures, agents and chance interact to bring about systemic change in IR), and learning the limits of what can't be known, the exercise is a knowledge-generating one.

<sup>42</sup> Berlin, Isaiah. *The Hedgehog and the Fox: An Essay on Tolstoy's View of History*. New York, NY: 1953: Simon and Schuster

<sup>43</sup> See Taleb, Nassim N. *The Black Swan*. New York, NY: Random House, 1997: 195 – 198. Taleb likens the process to tracing a water puddle to its original state as an ice cube. When looking at an ice cube on a table, we can confidently predict that it will turn into a puddle of water, and precisely compute the process. When looking at a puddle of water, even with the knowledge that this used to be an ice cube, the process of reconstructing the shape of the ice cube and its disintegration is far more difficult.