Psychology and aggregation in International Relations

Ross James Gildea
University of Oxford, UK

Abstract
Theories of decision-making grounded in political psychology have experienced a dramatic rise in the study of International Relations. There is widespread recognition of the benefits of incorporating insights from the behavioural sciences into analyses of political behaviour. However, some scholars have argued that the theoretical and empirical scope of these perspectives remains hampered by an unresolved issue: aggregation. While the fundamental unit of interest in psychology is the individual, most International Relations models concern patterns of collective decision-making in aggregate units such as states, bureaucracies, armed groups, transnational networks and institutions. This article contributes to the aggregation debate by providing a more optimistic portrait of its implications for interdisciplinary work. I argue that aggregation may be an overstated problem in International Relations and that a disciplinary preoccupation with it may hinder rather than pave the way for interdisciplinary theorizing.

Keywords
Political psychology, aggregation, International Relations, agent-structure problem, rationality, explanatory theory

Introduction
Theories of decision-making grounded in political psychology have experienced a dramatic rise in the study of International Relations (IR). Related research agendas, such as those associated with the affective ‘turn’ in IR, have also gained growing traction (Hall and Ross, 2015; Lopez and McDermott, 2012). Despite widespread recognition of the utility of incorporating insights from the behavioural sciences into analyses of political
behaviour, the theoretical and empirical scope of these perspectives in IR may be limited by an unresolved issue: aggregation. Although the difficulty of moving from models of decision-making at the individual level to higher levels of analysis is not confined to political psychology – being a feature of micro-foundational debates in IR theory more generally – the issue of aggregation is seen as a recurring challenge for psychological approaches to the examination of international politics.¹

For psychologists the fundamental theoretical unit of interest is the individual. This characteristic has made psychological insights readily compatible with the methodological individualism of fields such as microeconomics, contributing to the successful emergence of the sub-discipline of behavioural economics (Thaler, 2018). In contrast, incorporating psychological findings into IR may present a more difficult challenge. While theoretical and empirical work at the individual level may be quite straightforwardly adapted to analysis of the political behaviour of individual actors or the ‘first-image reversed’,² most IR models concern patterns of collective decision-making in aggregates such as states, bureaucracies, armed groups, institutions and transnational networks (Powell, 2017: S265). Much of these forms of social and political organization are designed to include checks and balances aimed at mitigating the effects of individual preferences. Decision-making processes in group and institutionalized settings can also be cumbersome and take time, providing further opportunities to counteract individual judgements and biases (Johnson, 2015: 760). How psychological mechanisms, which are primarily individually embodied, may operate and exercise influence within complex group and institutional environments remains a crucial and contested question.

The renewal of social scientific work which draws on psychological findings, indicative to some of a ‘behavioural revolution’, appears to be an auspicious time to revisit this challenge. Failure to resolve or plausibly circumvent problems of this kind has been viewed as an impediment to the broad diffusion of psychological insights to IR during previous waves of behaviourally informed research (Stein, 2017: S255).³ Some scholars predict that the way aggregation is addressed in coming years will affect whether interest in psychological approaches diminishes over time or if this research programme exerts a more systematic, integrated and lasting influence on the field (Hafner-Burton et al., 2017). This article contributes to the aggregation debate by providing a more optimistic portrait of its implications for interdisciplinary work.⁴ I argue that aggregation may be an overstated problem and that a disciplinary preoccupation with it in IR may hinder rather than pave the way for interdisciplinary theorizing.

The article proceeds in the following manner. In the first section ‘Problematizing the aggregation problem’, I provide a brief overview of existing analytical strategies to tackle the aggregation problem as well as highlighting potential criticisms of these approaches. The article then addresses each objection, positing that the supposed limiting effect of aggregation on the diffusion of psychological insights to IR may be overstated. The second section ‘Psychology as a complex base’ develops the latter argument more fully. It first examines the nature of psychology as a complex base for theory generation, one that is beset by individual-level factors that can move in multiple and contradictory directions. I argue that a theory of aggregation that rests on such a base, and that seeks to bridge multiple levels of analysis across different actor types, will likely suffer from weak foundational claims. In light of these challenges, I then suggest that the
theoretical pay-offs of a renewed focus on aggregation are unclear and that focusing on the construction of more complex models of aggregation may obstruct, rather than promote, the advancement of interdisciplinary scholarship. The third section ‘Complexity and interdisciplinary theory building’ connects the preceding discussion to a broader interdisciplinary debate regarding the combination of knowledge from different fields to construct increasingly complex theory. The article concludes with a brief discussion of its main claims.

Problematizing the aggregation problem

On the face of it, aggregation appears to present a serious challenge to psychological explanations in international politics. As Johnson (2015: 760) notes, ‘the whole point of government is to ensure multiple voices and checks and balances so that rational decisions can, in theory, persist despite individual preferences and biases’. Institutions are ordinarily designed to offset the expression of individual-level bias. This may occur through improved information gathering, representation of divergent views or the development of mechanisms of oversight and accountability (Saunders, 2017). The extent to which psychological traits may shape political outcomes at higher levels of analysis therefore appears open to question. The resurgence of behavioural research in IR has, in the eyes of many IR scholars, reopened debates on how the study of individual judgements and choices might inform explanations of collective decision-making in complex, institutionalized environments (Hafner-Burton et al., 2017: S5).

In order to address aggregation, scholars have typically employed two main analytical strategies. One ‘quasi-behavioural’ approach has been to treat aggregate actors as unitary and to draw on findings from cognitive social psychology to impute patterns of preference formation and decision-making to these actors (Powell, 2017). Much of IR theory situates aggregate entities as single ‘actors’ and attributes properties such as rationality, interests, identities and beliefs, which are drawn from our conceptions of personhood (Wendt, 2004: 289). Due to its focus on characteristics of human psychology, it is natural that a major strain of psychological IR would follow suit, retaining the unitary actor assumption but diverging from conventional work in respect to the kind of properties which were ascribed to aggregates. In this vein, models of interstate bargaining, conflict, deterrence and crisis dynamics have been refined through behavioural challenges to the ostensibly reductionist psychological assumptions of conventional theories (Janis, 1982; Jervis, 1976). The essays collected by Jervis, Lebow and Stein (1985) in Psychology and Deterrence are an instructive example of this approach. In this volume, the authors challenge the rationalist and deductive logic of deterrence theory, which appears to fit poorly with empirical cases. Instead, they propose an alternative model that accounts for the sequential behaviour of states that either challenge or defend the status quo. They situate deterrence as a fundamentally psychological relationship and differentiate between the psychological problems faced by challengers and defenders. By exploring the ‘emotions, perceptions, and calculations’ that inform decision-making, the authors provide an alternative perspective on the sources of state action which advanced our understanding of the dynamics of deterrence (Jervis et al., 1985).
The logic behind this approach is pragmatic and its potential analytical benefits are straightforward. Scholars make a research bet that linking behaviourally informed preferences with the unitary actor assumption will result in theoretical propositions that possess superior descriptive accuracy than standard models while also retaining parsimony (Powell, 2017: S270). Psychological theories of IR can thus, the logic goes, make productive use of behavioural insights without requiring a major trade-off with regard to tractability, a sacrifice commonly flagged by critics of analytically eclectic approaches (Sil and Katzenstein, 2010).

The second main strategy, favoured by the majority of political psychologists, has been to maintain focus on the individual level of analysis and to explain variation in international behaviour primarily through models of elite decision-making (Hafner-Burton et al., 2017: S18). This involves opening up the ‘black box’ of the state and other aggregates, distinguishing between individuals and small groups across structural environments and focusing explanations on outcomes associated with the distinct psychological characteristics of these actors (Sasley, 2011: 453). Proponents of this approach highlight the consistency between psychological insights gained at the individual level and the unit of analysis in their studies (Axelrod, 1976). A number of contemporary studies have developed this research agenda, emphasizing individual cognition in explanations for foreign policy behaviour. For instance, Gallagher and Allen (2014) posit that variation in the ‘Big Five’ personality traits of leaders can help explain both individual policy choices and recurring patterns of international conflict. Several case studies, such as Stein’s (1994) analysis of Soviet leader Mikhail Gorbachev and Dyson’s (2006) study of UK Prime Minister Tony Blair, examine how consequential international decisions may be affected by the dispositional attributes of elite politicians (see also: Schafer and Walker, 2006). Similarly, in this special issue, Sazak and Kadercan (2020) link interpersonal elite networks and concepts such as interpersonal trust to alliance formation at the state level.

Powell (2017) identifies potential flaws in both strategies. First, he argues that treating aggregates as unitary creates a mismatch between empirical findings in behaviour decision research and our analytical assumptions. While we may possess convincing evidence that individuals often do not conform to conventional rationalist assumptions, this ‘is not compelling evidence for or against treating groups as unitary actors with behavioral or standard preferences’ (Powell 2017: S266). Second, he suggests that focusing explanations on key individuals raises familiar obstacles to generating IR theory; it is very difficult to build ‘bottom-up’ theories that connect individual traits to aggregate-level outcomes (Powell, 2017: S266). Psychological factors may only be of explanatory value for IR when they are systematically integrated into theoretical frameworks ‘that take into account the structural, economic, and cultural conditions within which policy makers work’ (Goldgeier and Tetlock, 2001: 68). In light of these criticisms, Powell concludes that behavioural IR may have a modest impact on the field as a whole.

Despite the above misgivings, there is reason to believe that Powell’s conclusions are overly pessimistic and that the aggregation problem may be overstated. One potential source of optimism is the rise in volume of psychologically inspired studies across the discipline. The 2017 special issue of *International Organization*, devoted to *The Behavioral Revolution and International Relations*, is indicative of this rise. As Kertzer and Tingley (2018: 320) note in their review of the field, ‘the past decade and a half has
witnessed a surge of interest [in psychological approaches to IR], including in areas of the field that were historically less disposed to psychological work’. IR scholars are employing psychological insights in conjunction with diverse methodological tools, to investigate questions at multiple levels of analysis, across a growing array of research areas. Psychological inspiration is particularly well represented in research programs in international security, foreign policy, IR theory and human rights (Kertzer and Tingley, 2018: 322–324). If aggregation is having a severe, limiting effect on the diffusion of psychological insights, this does not appear to be reflected in the discipline.

Powell’s objection to the first analytical strategy for aggregation appears open to question. He contends that combining non-standard preferences with the unitary actor assumption will, by virtue of a mismatch between theoretical constructs in IR and empirical findings in behavioural research, impede the influence of psychological IR. However, as many of the most influential paradigms in IR theory rest on dubious micro-foundational assumptions, as well as simplified abstractions at higher levels of analysis, this does not seem to pose a major challenge to psychological approaches. For example, theories within the realist paradigm – including structural realism – have been shown to rely on unsustainable lower-level assumptions about human nature and psychology (Freyberg-Inan, 2004, 2006). In his Theory of International Politics, Kenneth Waltz (1979: 7) is explicit that his aim is simplification and ‘moving away’ from empirical referents. Moreover, Goldgeier and Tetlock (2001: 68) outline how the explanatory value of realist and constructivist theories is dependent on a ‘range of implicit psychological assumptions’, many of which have gone unremarked upon, let alone incorporated into a theory of aggregation. Because the influence of major IR paradigms has not relied on a tight deductive link between lower-level assumptions and theoretical predictions at higher levels of analysis, it is unclear why the success of theories that combine non-standard preferences with the unitary actor assumption would be limited by this criterion.

In addition, the permissibility of distinguishing theoretical constructs from empirical referents is not simply a matter of precedent but of the requirements of good theorizing. Substantial and transferable theoretical ideas typically rest on ‘. . .enabling assumptions about the world that necessarily elide and simplify so as to allow theorists to say useful things about it’, indicating that a tight link may be not only unnecessary but counterproductive (Levine and Barder, 2014: 868). This form of abstraction is an important step ‘. . .for a theory to move beyond ad hoc description’ (Gunitsky, 2019: 712). As Jackson (2011: 151) notes, this strategy is pervasive in IR, including in analyses of international organizations and networks (Hafner-Burton and Montgomery, 2006), security communities (Adler and Barnett, 1998) and political discourses (Hansen, 2006), among other research domains. Situating studies at the level of analysis appropriate to the research problem, and drawing on behavioural findings to construct new, if simplified, models of aggregate behaviour, therefore appears perfectly in tune with conventional IR theory building. This approach also allows – as illustrated by Jervis (1976) in his seminal work – behavioural insights to be fruitfully incorporated into existing frameworks.

Powell’s reservations about the second analytical strategy, that behavioural theories focused on individuals will face formidable obstacles in linking individual preferences to international outcomes, also may not be entirely justified. A growing strain of IR scholarship takes seriously the role of individual traits in affecting foreign policy outcomes
(Gallagher and Allen, 2014; Horowitz et al., 2015). Several studies have highlighted the considerable freedom leaders can have to make foreign policy decisions, in both autocratic and democratic states (Holmes, 2018; Saunders, 2011; Weeks, 2014). For example, Byman and Pollack (2001) engage with five historical case studies to show that ‘individuals play a central role in shaping international relations, including the causes of war, alliance patterns, and other areas that international relations scholars consider important’. For the authors, aggregation does not present a major challenge because it is attributes of the individual-as-actor that are the primary causal drivers of state behaviour. In their comparison of Iraq under Saddam Hussein and Syria under Hafiz al-Asad they find that, while the political and decisional environments faced by both leaders had marked similarities, differences in the leaders’ personal traits explain significant discrepancies in the foreign policy of both states, including matters of war and peace (Byman and Pollack, 2001: 128–132). This literature illustrates that, to the extent that leaders exercise discretion in decision-making, the aggregation problem may not arise in a meaningful way, as there will be consistency between psychological insights gained at the individual level and the unit of analysis.

Aggregation may be a more salient issue when choice outcomes are the product of collective decision-making. Recognizing this, IR scholars are paying increasing attention to patterns of collective decision-making in executive government, where groups rather than individuals are the unit of analysis. LeVeck and Narang (2017) conduct bargaining experiments to test the effects of individual preferences and relative group size on the outcomes of international crises. Mutz (2002) examines how the composition of social networks may shape what information actors consider when making decisions. Other studies have modelled dispositional traits of leaders with aspects of their decisional environment (Saunders, 2017). IR scholars are therefore already active on a promising research program that directly addresses different components of aggregation, including how groups deliberate, process information, interact with their institutional context and arrive at policy outcomes. This said, it is not necessarily the case that collective decision-making at the executive level actually poses a severe aggregation problem. Executives usually have rules and procedures when selecting policies, making patterns of aggregation easier to discern. In many democracies the chief executive has the final say, thereby helping to resolve the aggregation issue.

A further strand of literature has addressed the issue by directly theorizing collective behavioural phenomena. Work on emotions in IR has produced a lively discussion about how traits that are primarily individually embodied may occur at the group level. In contrast to theories that impute behavioural phenomena to aggregates under the unitary actor assumption, Hall and Ross (2015) suggest that traits such as emotion can acquire a genuinely shared dimension. Focusing on the influence of ‘affect’ in IR, that is the ‘embodied mental processes and the felt dimensions of human experience [that] influence thought and behaviour’, the authors suggest that affective experience may be jointly shared by individuals and groups in a top-down, bottom-up or horizontal fashion (Hall and Ross, 2015: 3). Similarly, Mercer (2014: 515) argues that ‘...group-level emotion is powerful, pervasive, and irreducible to individuals’ and that its distinctive properties can help us to better understand the behaviour of aggregate actors, including states. In his case study of the Rwandan genocide, McDoom (2012: 1) argues that psychology extends beyond the individual and that central emotions such as fear can ‘activate psychological processes at
the group level’. Although we may not fully grasp the mechanisms that give rise to and shape group emotion, this research program is an instructive example of how the challenge of aggregation may be overestimated.

In sum, when individual leaders make decisions largely in isolation, aggregation does not seem to be a meaningful problem. Neither does it appear to pose a particularly severe challenge in decisional units that follow clear procedures or are led by a powerful chief executive. In situations where group dynamics and institutional context are more consequential, IR scholars have already made inroads into modeling aggregation processes as well as directly theorizing behavioral phenomena at the group level.

One might therefore wonder if aggregation presents as significant a challenge as is sometimes claimed. In fact, aggregation seems to be something of a disciplinary preoccupation in IR. As Hafner-Burton et al. (2017: S13) acknowledge, its interest in aggregation distinguishes behavioral IR from other social science disciplines, many of which examine phenomena at the collective level without detailed specification of processes of aggregation. Sociology, for example, regularly deals with group behaviors such as public demonstrations and terrorism without expending much time on specifying the mechanisms of transmission from the individual to the collective (Rosental, 2013; Turk, 2004). For instance, Mark Granovetter’s (1978) influential work on threshold models of collective behavior, such as riots, analyzed divergent outcomes at the group level (e.g. whether a riot takes place or not) through the distribution of variance in individual preferences. Threshold models have also been applied to other forms of collective behavior, such as strikes, migration and voting, without setting out or testing particular mechanisms (Granovetter, 1978: 1423). In this work, a number of sociologists have drawn heavily on findings from psychology, as well as the behavioral sciences more broadly. Despite this, aggregation has not proved a major obstacle to interdisciplinary borrowing and theory construction to better our understanding of collective action.

Many economists have been similarly undeterred from incorporating behavioral insights into their models and situating their work across multiple levels of analysis. This interdisciplinary program is evident from the inception of modern macroeconomics. John Maynard Keynes’ (1936) classic work General Theory of Employment, Interest, and Money integrated psychological factors into explanations for macroeconomic processes and outcomes. Contemporary work by Nobel laureates George Akerlof and Robert Shiller (2010) built on Keynes’ legacy, connecting the psychological traits of individuals – theorized as more intuitive, emotional and irrational than standard models assume – to major financial events at the collective level, such as inflated house prices, failing capital markets and global financial crisis. Shiller (2005) also combines individual psychology with structural economic factors, such as technological change, to explain speculative bubbles. Although we cannot speak for the field as a whole, this literature illustrates how major economists have established a successful research program that relates individual-level characteristics to outcomes at the level of the national and global economy. They have done so without precise theorization and testing of mechanisms of aggregation, a feature that appears to have done little to restrict the appeal and influence of their theories.

The issue of aggregation in psychological IR seems no more challenging than that which has faced sociologists or economists. Linking individual psychology to the behavior of a government executive or an international institution is hardly more
complicated than connecting these traits to shifts in the global economy (perhaps much less so). It is therefore apt to question why the problem persists as a site of heated debate in IR. Our fellow social scientists have, by not getting bogged down in the specification of aggregation mechanisms, managed to create vibrant and influential research programs. Perhaps it is time that IR scholars followed suit. It may be our disciplinary preoccupation with aggregation, rather than the issue itself, which will prove a greater obstacle to interdisciplinary work.

**Psychology as a complex base**

This section develops the argument that focusing on the construction of models of aggregation may obstruct, rather than promote, the advancement of interdisciplinary scholarship. It does so by advancing two claims. First, I argue that psychology provides a complex base for generating a theory of aggregation, one that is beset by individual-level factors that can move in multiple and contradictory directions. A theory that rests on such a base, and that seeks to bridge multiple levels of analysis across different actor types, will likely suffer from weak foundational claims. Second, I posit that a research agenda focused on aggregation currently lacks clear theoretical pay-offs. It would also be at odds with much of theory building in the social sciences, which involves the formulation of simplified models to make inferences about more complex behaviour.

The adoption of psychological findings has had a profound effect on the study of decision-making in IR. Psychologists’ experiments revealed systematic patterns in individuals’ mental states, providing insights into how information processing, emotions, identity and beliefs impact preference formation. However, despite these gains, attempts to build a theory of aggregation on a psychological base are likely to encounter severe difficulties. One reason for this pertains to the applicability of experimental findings to collective decision-making in real-world settings. As Renshon and Kahneman (2017: 73) point out, behavioural decision research involves ‘theory of choices between gambles with specified probabilities and monetary outcomes’. Policy choices in the real world, relating to substantive scenarios such as the waging of war or negotiating peace, are much more complex than those analysed in psychological experiments. Outside of a laboratory, psychological mechanisms are shaped by an innumerable array of environmental factors, and tend to move in multiple and contradictory directions. Additionally, psychologists currently know very little about how competing or conflictual tendencies interact. Given that a theory of aggregation in behavioural IR would rest on a base of contingent and complex human behaviour, it is unclear on what, ultimately, such a theory would stake its foundational claims.

Consider prospect theory, a key part of the ‘heuristics and biases’ research program (Renshon and Kahneman, 2017: 64). Due to its emphasis on the role of both decision maker and environment in determining choice outcomes, prospect theory has proven flexible in its application to political behaviour, especially models of conflict (McDermott, 2004). However, the premium the theory places on concepts such as ‘loss’ and ‘value’ in actors’ calculations of risk raises obvious questions of contingency. In global politics, perceptions of risk and reward, as well as experiences of gains and losses, are rarely as neatly quantifiable as in the experiments devised by Kahneman and Tversky (1979,
In reality, perceptions of risk and value – whether social or material – arise in the context of complex relational identities. Group membership affects actors’ reasoning and assessments of their social reality, including their propensity to rely on heuristics or display bias (Miyamoto and Kitayama, 2002).

How and when prospect theory is relevant to decision-making is therefore dependent on variation in actor identities and the particular context within which preference formation takes place, including for members of the same decisional unit. In conflict scenarios, for example, individuals within the same group may place contrasting values on different features of the conflict (e.g. territory, military casualties, reputation). On the basis of distinct reference points, individuals’ perceptions of gains and losses, as well as their attitudes towards risk, may move in contradictory directions. It is unclear how a theory of aggregation might account for these kinds of contingent tendencies among individuals, how we might specify discrete mechanisms of intra-group transmission, and attribute variables appropriate at the individual level to higher-level units. Moreover, because the aggregation process may differ depending on the psychological mechanism and actor type under examination – it is plausible that governments, transnational networks, institutions and armed groups will exhibit distinct group dynamics – it is unclear whether attempts to build this kind of theory are viable.

Investigation of aggregation may place intense demands on the research process, with few clearly defined rewards. Following McCloskey (1956: 287), aggregation raises a host of related questions for the researcher, such as:

‘the general setting in which a decision is made; the actors’ perceptions of their roles and the actual authority they are free to exercise; the information, attitudes, and beliefs they bring to the decision-making event; the actors’ views of the alternative choices and consequences; the actual and perceived goals of the organization and its actors; the personality, skills, and needs of the decision-makers; and dozens of other matters of a similar order.’

Amid the many factors germane to aggregation, IR scholars currently lack a framework or set of guidelines for deciding which variables to focus on. It is also far from certain whether efforts to develop a theory of aggregation, engaging as it must with the above issues, will be matched by commensurate theoretical pay-offs. A theory that spans different actor types, psychological traits and decision-making contexts – one that would preclude the need to repeat intense investigative procedures in individual studies and across sub-disciplines – is precisely the type of theory that eludes political science. Moreover, it is hardly axiomatic that analysis of aggregation will yield important insights. To date, we possess no significant evidence showing that extant psychologically informed studies reached erroneous conclusions due to problems of aggregation. As McCloskey notes (1956: 292–293), ‘I have yet to be persuaded that an intensive study of the “networks of communication” within decisional units is likely to account for much of the difference between, say, decisions that lead to war and those that lead to peace’. A sizable gap exists between IR’s disciplinary preoccupation with aggregation and the identification of clear theoretical pay-offs that may follow from its investigation, as well as the formulation of working theories or guidelines that could provide a meaningful template for researchers.
to come to grips with it. Until this gap is bridged, it seems there is little to be gained from allowing aggregation to impede interdisciplinary work.

Given that IR scholars have already devised inventive ways to attribute individual-level characteristics to the collective level, shifting the focus to more complex models of aggregation would also be at odds with much of theory building in the social sciences. This typically involves the creation of simplified models to make sense of more complex patterns of individual and group behaviour. As Holmes (2018: 258) puts it in his interdisciplinary study of neuroscience and diplomacy,

‘while the extrapolation from simple designs to more complicated interaction may not be perfect, and perhaps often is messy, it is a justifiable jump as it represents the core method that neuroscience, and any other science including the social, uses to make predictions about complex interactions. Theory building requires the very type of simplified models that neuroscience, and other reductionist approaches, provide.’

Further, the idea that failing to adequately theorize aggregation impeded the diffusion of previous waves of behavioural research – a negative harbinger for today’s behavioural revolution – appears contestable. Rather, the reverse may be true; commitment to building stylized models, without laborious specification of the processes of aggregation, may well have been a central feature of their success. Jervis’ (1976) landmark text on perception in international politics is an instructive case.

Jervis identifies sites of divergence between perceived and actual state intentions on the part of decision makers, relying on an ‘error-and-bias portrait of the foreign policy maker’ (Tetlock and Goldgeier, 2001: 68). Crucially, Jervis (1976: 8) narrows his analysis by eschewing intercultural differences in perception between individuals and groups, instead looking to patterns of cognition that are collectively shared. He argues that ‘theory and explanation need not fill in all the links between cause and effect [. . .] one can always ask for the links between links. High-density theories have no privileged status; they are not automatically illuminating or fruitful’ (Jervis, 1976: 13). It is difficult to maintain that Jervis’ decision to elide aggregation hindered the interdisciplinary process or the book’s pervasive influence. Nor does his analytical strategy appear to have faced a bigger obstacle than that which would have been posed by a need to specify detailed mechanisms of aggregation. By simplifying his view of human cognition, interpersonal interaction and decisional context, Jervis lifted barriers to interdisciplinary borrowing and enhanced our theoretical understanding of decision-making in international politics.

Following Renshon and Kahneman (2017: 73), extrapolating from psychological theories of choice – involving clearly determined probabilities and monetary outcomes – to complex social behaviour is extremely difficult, as ‘there is too much else going on’. The complexity of psychological mechanisms provides a weak base for a theory of aggregation to stake its foundational claims. While Powell is correct in identifying a mismatch between empirical findings in behavioural research and some of the analytical assumptions of psychological IR, it is less obvious that focusing increased theoretical and empirical attention in this area would prove fruitful. There seems to be a sizable gap between IR’s disciplinary preoccupation with aggregation and the identification of clear theoretical pay-offs that may follow from its investigation. IR scholars lack a set
of working theories or guidelines to begin the endeavour. Moreover, the emphasis on aggregation is at odds with much of theory building in IR and – Jervis’ work is a prime example – the impact of psychological studies on IR has hardly been severely impeded by the issue. IR scholars may be better advised to follow the example of the sociologists and macroeconomists, to avoid a ‘domestication dynamic’ – that is, reconstruing insights from outside the field to fit with the parameters of current discussions taking place in IR – and to ensure that aggregation does not act as a barrier to productive interdisciplinary exchanges with the behavioural sciences.

Complexity and interdisciplinary theory building

In interdisciplinary studies, what is being borrowed – concept, theory or methodology – and to what end, presents distinct analytical challenges. In terms of theory building, the discussion in the preceding section raises a question that speaks to the broader ambitions of interdisciplinary IR: when is it desirable to combine knowledge from different fields to construct increasingly complex theory? While a definitive answer to this question remains elusive, I argue that adding complexity should be conditional on at least three considerations: the theoretical goals of the researcher, the complexity of the political event or process under examination, and our current level of knowledge of it.

A central consideration in whether to add complexity to theory is its underlying purpose. Scholars may be motivated by particular theoretical goals, including realism, precision and generality. Each of these infers a different approach to representing and understanding the world, including distinct perspectives on complexity (Godfrey-Smith, 2006: 725). For instance, one approach to IR theory is to sacrifice elements of realism and precision to construct generalizable models, enabling comparison and causal inference across multiple cases. This approach generalizes by relying on abstraction, that is ‘new ideas or conceptions are formed by considering several objects or ideas and omitting the features that distinguish them’ (Rosen, 2014). Representational accuracy is sacrificed, but with the objective of identifying salient features of the world. There are no categorical rules on how to make this trade-off. The aim for researchers is to decide ‘where distortion is least dysfunctional and where such accuracy is absolutely essential’ (Singer, 1961: 79). Where IR scholars judge knowledge produced in other fields – such as the behavioural sciences – to illuminate key features of the political world, combining interdisciplinary knowledge to construct more complex theory seems permissible.10

Where scholars are concerned with enhanced realism and precision, drawing on interdisciplinary insights to form more complex theory may be warranted. In particular, IR theory that aims at prediction or policy relevance stands to benefit from added complexity and interdisciplinary borrowing. Consider predictions in the social sciences. In his comprehensive study of forecasting, Tetlock (2005) outlines how the trade-offs associated with simplified models often fail to contend with the complexity of the political world. Tetlock’s work shows that ‘grossly inaccurate forecasts are more common when experts employ a single parsimonious approach and rely excessively upon broad abstractions’ (Sil and Katzenstein, 2010: 12). By contrast, theories that cross the boundaries of disciplines and paradigms, incorporating diverse forms of knowledge, generally have more predictive power (Tetlock, 2005: 88–91). The same is true of policy-orientated
research. Focus on stylized models has contributed to a ‘flight from reality’ in the social sciences, where scholars ‘have lost sight of what they claim is the object of their study’ (Shapiro, 2005: 2). Its extent may be debatable, but many agree on the existence of a gap between much of IR theorizing and the demands of policymakers. A countervailing tendency to ‘re-complexify’ theory may help to address this gap. In line with Sil and Katzenstein (2010: 19), researchers can combine knowledge to ‘reflect, rather than simplify, the complexity and multidimensionality of social phenomena of interest to both scholars and practitioners’. Efforts in this direction could produce work that is both academically and practically useful, an outcome that would seem to justify more complex and interdisciplinary theory.

Extending from this, more complex, interdisciplinary theorizing may be desirable in proportion to the complexity of the process or event under examination, given the utility of insights from other fields in highlighting its key features. Interdisciplinary research and conventional strategies for IR theory building tend to differ in their respective ontologies and attitudes towards complexity. Central to this is the principle of parsimony. While there is disagreement over its relevance and application, ‘parsimony remains a key concept in international relations theory’ (Gunitsky, 2019: 707). Among its variants, perhaps the most influential justification for parsimony is epistemological. As Gunitsky (2019: 707) puts it,

> the epistemological justification makes no claims about elegance or the nature of existence and instead uses parsimony as a way to produce generalizable explanations and improve causal inference [. . .] It rests on the idea that explanation requires generalization, and generalization requires abstraction.

On the other hand, interdisciplinarity begins from a different logic; it *does* tend to make claims about the nature of its subject, in that it is seen as a means of making complex phenomena more intelligible; ‘most scholars seem to agree that what justifies the production of integrative, interdisciplinary research is the complexity of reality’ (Holland, 2014: 11). Despite this difference, parsimony does not demand that a theory be maximally simple. Nor do we need to reject it to add complexity to theory. The epistemological variant does not aim for simplicity for its own sake but to highlight salient features of the world to advance generalizable causal explanations. After all, ‘the theory should be just as complicated as all our evidence suggests’ (King et al., 1994: 20). Therefore, where the nature of the subject under examination is sufficiently complex, the construction of good theory may require added complexity, including insights garnered from other fields. Scholars may need to look beyond their own discipline to gather tools and knowledge that help take account of its underlying complexity, and to give expression to its most salient characteristics.

Another criterion might be our current level of knowledge of the political event or process under examination. It is best not to pursue an elegant, stylized theory until we have sufficient knowledge of the subject matter. As King et al. (1994) write, ‘we do not advise researchers to seek parsimony as an essential good, since there seems little reason to adopt it unless we already know a lot about a subject’. When we already know a lot about a topic, creating parsimonious models is more manageable and may help avoid unwieldy theories.
However, if a topic has received little theoretical and empirical attention, attempts to construct simplified models may not prove fruitful, as we would lack knowledge of the main features that allow for parsimonious theorizing. In this scenario, scholars may find it useful to combine interdisciplinary knowledge to guide their preliminary investigations. Interdisciplinary research programs in IR, like other areas of the social sciences, should ‘develop theories that aim for a fruitful combination of simplicity and strength’ (Healy, 2017: 119). By widening its analytical net, interdisciplinary scholarship can help us identify the salient features of a subject and begin the process of theory building. Rather than being pursued at the expense of parsimony, an approach that is open to complexity at the outset of the research process will be conducive to constructing simplified theories later on, once we have accumulated adequate knowledge to do so. This approach is consistent with a descriptively rich and theoretically rigorous discipline (Besbris and Khan, 2017: 147).

What does this mean for the study of aggregation? With regard to the first criterion, the theoretical goal of the researcher, the implications are mixed. If representational accuracy is sacrificed to aid generalizability, there appears to be few incentives to investigate aggregation. As highlighted in the previous section, we currently do not have compelling evidence that aggregation is a critical site of inquiry for psychological IR, or for other areas of the social sciences. Until more convincing evidence is forthcoming, adding complexity by accounting for aggregation appears somewhat superfluous. However, where the theoretical motivation is one of realism and precision, and representative accuracy is paramount, aggregation could prove to be a worthwhile area of study. It is plausible that aggregation processes have a meaningful effect on decision-making outcomes in particular contexts. Increased attention to it may therefore improve predictions or policy lessons in those specific domains, albeit at the expense of building the kind of transferable theory of most interest to the discipline. Moreover, making a research bet on aggregation may be justified when the subject of investigation is complex, or we have little preliminary theoretical or empirical knowledge of it. In such situations, the causal implications of how attributes of units translate across levels cannot be dismissed out of hand, particularly if the underlying complexity of a process or event relates to different levels of analysis. Casting a wide net in preliminary investigations, inclusive of issues of aggregation, could strengthen the empirical foundation of subsequent theories, even if aggregation is ultimately deemed of marginal theoretical importance.

**Conclusion**

The resurgence in IR work grounded in political psychology has reopened debates surrounding aggregation. A number of scholars have argued that, until the issue is satisfactorily resolved, it will remain an obstacle to interdisciplinary theorizing and may ultimately limit the impact of psychological approaches on the field. This article has sought to advance the debate on aggregation, providing a more optimistic portrait of its implications for interdisciplinary work. I argue that the challenges it poses may be overstated, and that a disciplinary preoccupation with aggregation in IR may hinder rather than pave the way for interdisciplinary theorizing.

The article illustrates how existing approaches to aggregation are consistent with conventional strategies for theory building. Scholars have already devised innovative ways
to tackle the problem and it is on this basis that previous waves of research in psychological IR have had a significant impact on the field. Given the current proliferation of studies in this domain, it seems contradictory to talk of a ‘behavioural revolution’ sweeping the discipline while also citing aggregation as a major obstacle to precisely this sort of influence. Aggregation has not prevented the diffusion of behavioural research, nor does it seem to be the stumbling block that it is often considered to be.

Moreover, the article considers how other social science disciplines, such as sociology and economics, have used psychological insights to develop macro-level theories without laborious specification of mechanisms of transmission from the individual to the collective. Concern about aggregation appears to be a disciplinary preoccupation in IR, with the potential to inhibit interdisciplinary work. It is not axiomatic that devoting more time and resources to the investigation of aggregation will be matched with commensurate theoretical pay-offs. Even if scholars pursued this line of research, any prospective theory of aggregation based on behavioural findings may prove troublesome. Psychology provides a complex base for theory generation, one that is beset by individual-level factors that can move in multiple and contradictory directions. A theory that rests on such a base – seeking to bridge multiple levels of analysis across different actor types – will likely suffer from weak foundational claims.

One might argue that psychology has formed a key part of the discipline since the inception of IR, a connection likely to continue in the future. Research in psychology has aided scholars in their efforts to build theories, identify scope conditions, refine concepts and interrogate assumptions (Holmes, 2018: 256). Future work can look to successful models of interdisciplinary borrowing that have enriched the discipline. One such example is the adaptation of prospect theory to IR. Prospect theory has been shown to possess numerous analytical benefits that have helped to move the discipline forward. Gains include providing an empirically grounded alternative to rational choice theory, incorporating dynamic change into models, highlighting the role of context and loss in decision-making and opening up new pathways of research (McDermott, 2004: 290). Again, these benefits were garnered without excessive focus on mechanisms connecting individual traits to higher levels of analysis.

Moving forward, not everyone will make the same research bets on the future of psychological IR. As I have acknowledged, there may be occasions when it is desirable to re-complexify IR theory, and to engage with aggregation more directly, especially in research domains that are themselves complex, we possess little knowledge of, or that are relevant to policymaking. This being said, we should not exaggerate the analytical challenges posed by aggregation, nor should we subject psychological insights to a domestication dynamic that would preclude the kind of interdisciplinary theorizing that continues to enrich the discipline.

**Funding**

The author received financial support for this research from the Economic and Social Research Council and the National University of Ireland Travelling Studentship in Economics, Politics, and Social Policy.
Notes

1. I adopt the term psychological IR to refer to interdisciplinary studies that draw on conceptual, theoretical and methodological insights from psychology, as well as the behavioural sciences more broadly. The term is used interchangeably with behavioural IR.

2. First-image reversed refers to a research programme which, rather than investigating how individuals affect international politics, analyses how political phenomena at higher levels of analysis impact upon individuals (Kertzer and Tingley, 2018: 12).

3. Under this term I include evolutionary, neurobiological and affective approaches.

4. There is disagreement over the disciplinary status of IR, with scholars variously contending that it is a subfield of political science, a discipline in its own right or an inherently hybrid field that encompasses politics, economics, culture and history, among other forms of human activity and inquiry. Following Turton (2015: 245), I conceive of the academic discipline as emerging from ‘. . .the institutions, discourses, professionalization and the identity of academics who self-identify as belonging to a given discipline’, as well as a general, collective understanding of its members’ intellectual focus. I take interdisciplinary research in IR to refer to the exchange, borrowing and integration of insights from distinct spheres of academic pursuit and their incorporation into IR studies to address questions of import to the field. On the debate surrounding the disciplinary status of IR, and contested attitudes towards interdisciplinary work in the field, see Grenier et al. (2015); Yetiv and James (2016); Wæver (2007); Holland (2014).

5. Multilevel work in the civil wars literature, such as the study of identity and group cohesion, also appears to be influenced by psychology, without adopting the formal vernacular of the discipline (Kertzer and Tingley, 2018: 322).

6. For a discussion of this, see Gunitsky (2019: 771).

7. Jervis examined how cognitive limitations impact rational decision-making within a realist framework.

8. Prospect theory is a framework for analysing risk and uncertainty in decision-making. It suggests that people judge utility against a reference point, weigh losses more heavily than gains and are more prone to risky decisions having suffered a loss; see Kahneman and Tversky (1979).

9. For further discussion of these goals, see Godfrey-Smith (2006).

10. It is not axiomatic that combining interdisciplinary knowledge will lead to more complex theory. While that may be the case, insights from other disciplines may also be used to replace components of existing theory, or to make a theory more manageable and parsimonious.

References

Adler E and Barnett M (1998) Security Communities. Cambridge: Cambridge University Press.
Akerlof G and Shiller R (2010) Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism. Princeton, NJ: Princeton University Press.
Axelrod R (1976) Structure of Decision: The Cognitive Maps of Political Elites. Princeton, NJ: Princeton University Press.
Besbris M and Khan S (2017) Less theory: more description. Sociological Theory 35(2): 147–153.
Byman DL and Pollack KM (2001) Let us now praise great men: bringing the statesman back in. International Security 25(4): 107–146.
Dyson SB (2006) Personality and foreign policy: Tony Blair’s Iraq decisions. *Foreign Policy Analysis* 2(3): 289–306.

Freyberg-Inan A (2004) *What Moves Man: The Realist Theory of International Relations and Its Judgment of Human Nature*. Albany, NY: State University of New York Press.

Freyberg-Inan A (2006) Rational Paranoia and enlightened machismo: the strange psychological foundations of realism. *Journal of International Relations and Development* 9(3): 247–268.

Gallagher ME and Allen SH (2014) Presidential personality: not just a nuisance. *Foreign Policy Analysis* 10(1): 1–21.

Godfrey-Smith P (2006) The strategy of model based science. *Biology and Philosophy* 21(5): 725–40.

Goldgeier J and Tetlock P (2001) Psychology and international relations theory. *Annual Review of Political Science* 4: 67–92.

Grenier F, Turton HL and Beaulieu-Brossard P (2015) The struggle over the identity of IR: what is at stake in the disciplinary debate within and beyond academia? *International Relations* 29(2): 242–244.

Gunitsky S (2019) Rival visions of parsimony. *International Studies Quarterly* 63(3): 707–716.

Hafner-Burton EM, Haggard S, Lake DA, et al. (2017) The behavioral revolution and international relations. *International Organization* 71(S1): S1–S31.

Hafner-Burton EM and Montgomery AH (2006) Power positions: international organizations, social networks, and conflict. *Journal of Conflict Resolution* 50(1): 3–27.

Hall TH and Ross AAG (2015) Affective politics after 9/11. *International Organization* 69(4): 847–879.

Hansen L (2006) *Security as Practice: Discourse Analysis and the Bosnian War*. London: Routledge.

Healy K (2017) Fuck nuance. *Sociological Theory* 35(2): 118–27.

Holland D (2014) *Integrating Knowledge Through Interdisciplinary Research: Problems of Theory and Practice*. Abingdon, UK: Routledge.

Holmes M (2018) *Face-to-Face Diplomacy: Social Neuroscience and International Relations*. Cambridge, UK: Cambridge University Press.

Horowitz MC, Stam AC and Ellis CM (2015) *Why Leaders Fight*. New York, NY: Cambridge University Press.

Jackson PT (2011) *The Conduct of Inquiry in International Relations*. New York: Routledge.

Janis IL (1982) *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes, Second Edition*. Boston, MA: Houghton Mifflin.

Jervis R (1976) *Perception and Misperception in International Politics*. Princeton, NJ: Princeton University Press.

Jervis R, Lebow RN and Stein JG (1985) *Psychology and Deterrence*. Baltimore, MD: Johns Hopkins University Press.

Johnson DDP (2015) Survival of the disciplines: is international relations fit for the new millennium? *Millennium: Journal of International Studies* 43(2): 749–763.

Kahneman D and Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47(2): 263–291.

Kahneman D and Tversky A (1984) Choices, values, and frames. *American Psychologist* 39(4): 341–350.

Kertzer J and Tingley D (2018) Political psychology in international relations: beyond the paradigms. *Annual Review of Political Science* 21: 1–23.

Keynes JM (1936) *The General Theory of Employment, Interest and Money*. New York: Macmillan.

King G, Keohane RO and Verba S (1994) *Designing Social Inquiry*. Princeton, NJ: Princeton University Press.
LeVeck B and Narang N (2017) The democratic peace and the wisdom of crowds. International Studies Quarterly 61(4): 867–880.

Levine DJ and Barder AD (2014) The closing of the American Mind: ‘American School’ international relations and the state of grand theory. European Journal of International Relations 20(4): 863–88.

Lopez AC and McDermott R (2012) Adaptation, heritability, and the emergence of evolutionary political science. Political Psychology 33(3): 343–362.

McDermott R (2004) Prospect theory in political science: gains and losses from the first decade. Political Psychology 25(2): 289–312.

Miyamoto Y and Kitayama S (2002) Cultural variation in correspondence bias: the critical role of attitude diagnosticity of socially constrained behavior. Journal of Personality and Social Psychology 83(5): 1239–1248.

McCloskey H (1956) Review: concerning strategies for a science of international politics. World Politics 2(8): 281–295.

McDoom OS (2012) The psychology of threat in intergroup conflict: emotions, rationality, and opportunity in the Rwandan Genocide. International Security 37(2): 119–155.

Mercer J (2014) Feeling like a state: social emotion and identity. Institutional Theory 6(3): 515–535.

Mutz DC (2002) Cross-cutting social networks. American Political Science Review 96(1): 111.

Powell R (2017) Research bets and behavioral IR, International Organization 71(S1): S265–S277.

Renshon J and Kahneman D (2017) Hawkish biases and the interdisciplinary study of conflict decision-making. In: Yetiv SA and James P (eds) Advancing Interdisciplinary Approaches to International Relations. Cham, Switzerland: Palgrave Macmillan, 51–82.

Rosen G (2014) Abstract objects. In: Zalta EN (ed.) The Stanford Encyclopedia of Philosophy. Palo Alto, CA: Stanford University.

Rosental C (2013) Towards a sociology of public demonstrations. Sociological Theory 31(4): 343–365.

Saalschütz BE (2011) Theorizing states’ emotions. International Studies Review 13(3): 452–476.

Saunders E (2011) Leaders at War: How Presidents Shape Military Interventions. Ithaca, NY: Cornell University Press.

Saunders E (2017) No substitute for experience: presidents, advisers, and information in group decision making. The Behavioral Revolution and International Relations, Supplement, International Organization 71(S1): S219–S247.

Schaefer M and Walker SG (2006) Democratic leaders and the democratic peace: the operational codes of Tony Blair and Bill Clinton. International Studies Quarterly 50(3): 561–583.

Shaffer H (2005) The Flight From Reality in the Human Sciences. Princeton, NJ: Princeton University Press.

Shiller R (2005) Irrational Exuberance. Princeton NJ: Princeton University Press.

Sil R and Katzenstein PJ (2010) Analytic eclecticism in the study of world politics: reconfiguring and mechanisms across research traditions. Perspectives on Politics 8(2): 411–431.

Singer JD (1969) The level-of-analysis problem in international relations. In: Rosenau JN (ed.), International Politics and Foreign Policy: A Reader in Research and Theory. New York: The Free Press, 20–29.

Stein JG (2017) The micro-foundations of international relations theory: psychology and behavioral economics. In: The Behavioral Revolution and International Relations, supplement, International Organization 71(S1): S249–S263.

Thaler RH (2018) From cashews to nudges: the evolution of behavioural economics. American Economic Review 108(6): 1265–1287.

Tetlock P (2005) Expert Political Judgment: How Good Is It? How Can We Know? Princeton, NJ: Princeton University Press.
Turk AT (2004) Sociology of terrorism. *Annual Review of Sociology* 30(1): 271–286.
Turton LH (2015) The importance of re-affirming IR’s disciplinary status. *International Relations* 29(2): 244–250.
Waltz K (1979) *Theory of International Politics*. New York: McGraw-Hill.
Wæver O (2007) Still a discipline after all these debates? In: Dunne T, Kurki M and Smith S (eds) *International Relations Theories: Discipline and Diversity*. Oxford: Oxford University Press, 290–291.
Weeks JL (2014) *Dictators at War and Peace*. Ithaca, NY: Cornell University Press.
Wendt A (2004) The state as person in international relations theory. *Review of International Studies* 30(2): 289–316.
Yetiv SA and James P (2017) *Advancing Interdisciplinary Approaches to International Relations*. New York: Palgrave Macmillan.

**Author biography**

**Ross James Gildea** is a DPhil Candidate in International Relations at the Department of Politics and International Relations, University of Oxford. He is a tutor in political psychology and international relations at Mansfield College. His current research focuses on issue prioritization in transnational advocacy networks. He can be reached at ross.gildea@politics.ox.ac.uk.