Understanding (Mis)Behaviour and Bias: An Integrative Review of the Literature on Behavioural International Relations

By

Ciaran Christopher Burks

Submitted in fulfilment of the requirements for the degree:

Master of International Relations

Department of Political Sciences

Faculty of Humanities

University of Pretoria

Supervisors: Dr Robin Bourgeois and Dr John Kotsopoulos

Acknowledgements

For my mom, dad and brother, who've never left my side.

For my stalwart friends,

And for my Love, for lending me your support, your ears and your heart

Thank you all.

Declaration

I declare that this dissertation: *Understanding (Mis)Behaviour and Bias: A Review of the Literature on Behavioural International Relations,* is my own work. It has never before been submitted for any degree or examination in any other university. All the sources that I have used or quoted have been indicated and acknowledged as complete references.

Signed:

Ciaran Christopher Burks September 2018

Abstract

This study presents an integrative review of the literature on Behavioural International Relations (BIR). It undergoes a detailed analysis of six texts dealing with the core issues in the field BIR. These texts are used as a foundation from which the review: i) Creates a conceptual framework for the notion of "cognitive bias" and identifies the potential value of better understanding its role in International Relations (IR), ii) Identifies behaviour incongruent with traditional IR theory and iii) Shows that this incongruent or anomalous behaviour may in fact be described by theories already present in the IR literature, namely within the literature of BIR.

The eclectic sample this study uses is selected according to criteria clearly stated in the methodology. The review will briefly note the roots of a 'behavioural revolution' within the political sciences, the subsequent rebellion against it, its more recent re-emergence and the utility of this re-emergence for understanding certain puzzling phenomena in IR.

The review begins by defining terms and outlining exactly what these key terms mean in the context of this review. The review briefly demonstrates some of the failings of traditional IR approaches to explain certain phenomena in IR. This literature review makes use of illustrative cases throughout, in order to demonstrate that the approaches synthesised in this review are better at explaining certain phenomena than traditional IR theory. The review discusses the possibility that outcomes labelled as bizarre may in fact be fairly congruent with a different kind of political analysis – that is, the type of analysis that BIR can offer.

By illustrating that human decision makers (individuals and groups) act in some predictably biased ways – and attempting to explain why – this paper explores the possibility of more correctly specifying what we classify as behaviour or misbehaviour. The study shows that

events that have been considered the result of “misbehaviour” might simply be biased human cognitive processes playing themselves out. This review attempts to demonstrate the opportunity for more accurately defining and explaining baseline behaviour and enhancing our understanding of the causes, nature and results of cognitive bias in IR.

List of Abbreviations and Acronyms

AB – Association-based

BIR – Behavioural International Relations

IPE – International Political Economy

IR – International Relations

ISQ – International Studies Quarterly

NASA – National Aeronautics and Space Administration

NGO – Non-Governmental Organisation

PS – Psychophysical-based

SB – Strategy-based

TRA – Theory of Reasoned Action

TPB – Theory of Planned Behaviour

UK – United Kingdom

UN – United Nations

Table of Contents

Chapter 1

Introduction

1. On Structure

Chapter one of this study presents the outline of the integrative review which will take place. It identifies the research theme, presents the research question, outlines the goals and structure of the following review, outlines the methodology used in the review and presents the research structure of the paper. This chapter includes definitions of key terms for the purposes of clarity. Chapter one will give a very brief account of the roots of behavioural social sciences and its diffusion into International Relations. It discusses the methodological and theoretical foundations of Behaviouralism, and how these foundations inform the subfield of Behavioural International Relations.

Chapter two will create the basis for the following discussion of the pertinence of the field, and a framework from which we can attempt to answer the main research question. This section will aim to create a conceptual framework for the idea of cognitive bias. It will analyse the phenomenon in detail, drawing on a wide range of research from different fields. It will specify what it is, illustrate its importance for understanding IR phenomena and explore the ontological process present in BIR that finds cognitive bias at its epicentre.

Chapter three will embark on a review. It will review the methods, findings and conceptual frameworks of the six important texts. It will provide an exposition of BIR and break it down, in detail, into its constituent components, aims, findings and possibilities. It will show that there is a theory within IR that may account for apparently bizarre behaviour by relevant agents – the same phenomena discussed in the previous chapter. It will connect the previous chapter to applications and outcomes, demonstrating clearly that cognitive bias has a prominent role to play in remedying some of the mysteries in IR behaviour. This chapter helps to justify my argument that BIR and the concept of cognitive bias solves some major IR puzzles in a convincing way.

Chapter four shows that there are gaps and challenges to IR theory. It will show that some of these gaps exist because of IR theory's failure to regard cognitive bias as an important variable in its analysis. It shows that there exist phenomena in IR that are poorly explained by dominant paradigms. These illustrative examples will make it clear that current models of behaviour in IR are insufficient to deal with some puzzles. Further, by analysing these illustrative examples through a BIR lens, it will show the scope, relevance and importance of cognitive bias in creating a useful and insightful picture of the international world. This chapter argues that BIR – and the inseparable concept of cognitive bias – could be the appropriate solution to some of the problems of relevancy and accuracy of the dominant IR theories.

Chapter five will attempt to be introspective. It will, in a spirit of academic humility, analyse its own core arguments and investigation to determine i) if BIR and the associated concept of cognitive bias does indeed contribute to a better understanding of previously assumed "anomalous" events in IR, ii) the extent of this contribution and its limits or boundaries. This chapter will also highlight further gaps the study revealed and recommend those areas for further research. It also attempts to synthesise the key findings of the report and highlight the weight of evidence with justifies the position of this study.

2. Research Theme and Research Problem

The units of study in political sciences – individuals, states, political elites and systems of governments – are phenomena in flux. Objects of political study both influence and are influenced by the people that surround and study them (Berkenpas, 2016). The fluctuating nature of these units of analysis necessitates a simultaneous (re)evaluation of existing theories and practice in order to keep up with these complex actors and outcomes. The lack of theoretical uniformity in political science provides both challenges and opportunities. The principal theoretic approaches in International Relations (IR) remain largely polarised, especially when one examines them in broad terms (Slaughter, 2011). The theoretical approach of BIR presents an opportunity to consolidate this polarisation. It promises to build a useful analytical bridge between traditional theories in IR, using a multi-disciplinary approach (James, 2007).

Can we – using International Relations theory – adequately explain apparent lapses in rational (strategically utility maximising) decision-making and the failure of policymakers and leaders to weigh relatively clear costs and benefits and make predictable decisions? I think we presume too much if we answer 'yes'.

International Relations theory has a tradition in the rationalist school of thought. Standard theories, positing states as actors in the international arena, are the norm for International Relations scholars. Such schools include the realist, institutionalist, liberal, constructivist, English and critical schools. These schools differ, essentially, in the variables that they emphasise when studying state decision-making. They include: ideological stance, military power or material gain (Slaughter, 2011). Rationalist approaches to explaining and predicting the actions of states focus heavily on principles of benefit, cost and expected utility. Such rationalist theory often fails to acknowledge certain pertinent realities of the human decision-making process, namely, that these processes are often subconscious and irrational (James, 2007).

The review will note and give clear examples of the lack of theoretical explanation for some international relations phenomena. It will then show that the theoretical approach of BIR can help to address some of these failings and consolidate IR theory to make it a more powerful explanatory tool for scholars and policy makers.

Disturbingly, it seems that, given the evidence, political specialists are as good at predicting medium and long-term political outcomes as political novices. That is to say, neither are very good at predicting the future (Tetlock, 2005). Other studies by Hafner-Burton, et al. (2013 & 2017), Mintz & DeRouen Jr (2010), Stein (2017) and many others indicate the same problems with forecasting political outcomes. A counterpoint argues that perhaps it is not the place of political scientists and political commentators to predict future outcomes. I would agree that it is not the job of a political expert to be a prophet, given the randomness and complexity of the real world. I would, however, argue that – inasmuch as their predictions are based on accurate theory – we would expect their forecasts to outperform the politically ignorant. It is plausible that the lack of any longer-term forecasting ability – even a very limited one – is a symptom of assumptions about the world and actors that should be updated.

BIR aims to analyse interactions not only between states, but between the individuals that

represent them. In so doing, it attempts to increase its own explanatory power as well as make predictions more robust, since individual relationships affect international relations significantly. Inserting individual considerations into explanatory and predictive models, or making them more important variables, may make them more accurate.

IR research is still largely embedded in traditional theories, with more than 4 out of 5 IR publications consisting of some form of paradigmatic analysis and with 75% of the time in introductory courses devoted to traditional IR paradigms (Hellmann, 2011).

However, none of these traditional paradigms consider cognitive bias a central variable in their analysis of IR. This is, potentially, evidence of a gap in the theoretical literature. There is, however, a growing interest in experimental evidence that is casting light on the human decision-making process in complex, high-risk and uncertain environments. Political elites are consistently forced to make choices in just such environments, with significant impacts on the international arena. This may explain some of the anomalous behaviour we must grapple with in IR.

By way of summary then, the research problem is this: human beings have been shown to fall short of rational-model assumptions. They make predictable errors in judgement based on cognitive biases. However, the most popular theoretical lenses in International Relations have not adapted their models for this reality, and thus misunderstand behaviour that is – promisingly – understandable, albeit with a different theoretical lens. By using BIR as this lens, it might be possible to modify our collective understanding of events in International Relations, which would contribute to establishing and maintaining peace, stability and cooperation in the international order.

By making use of the concepts and methodological frameworks present in the BIR literature, it may be possible to understand phenomena in IR that are not satisfactorily explained by traditional IR theory. BIR shows promise in shedding light on the decision-making of political elites, on cooperation, compliance and institutional design. The behavioural revolution that transformed the social sciences long ago did not truly diffuse into IR, but it might yet, and the important insights, empirical strategies and experimental research introduced by this revolution will be crucial in informing the research agenda and improving the relevance of IR scholarship.

3. Research Question and Hypothesis

This integrative review will make use of illustrative examples throughout in order to show – tangibly – that there is indeed a strong argument for rethinking how we, as scholars of IR, analyse (ir)rationality and the (mis)behaviour of significant actors in the field, including both states and individuals. In order to complete this task, the analysis will aim to answer the main question:

In light of "anomalous" behaviour by actors in IR, often reduced to irrationality by traditional IR theory, might BIR – which focuses on cognitive bias as a core variable in its analysis – better explain such behaviour?

Related sub questions include:

i) What is cognitive bias and how is it conceptualised in the literature?

ii) What kind of events does traditional IR theory struggle to adequately explain?

iii) How does BIR deal with these events?

And

iv) Does BIR provide a compelling framework for approaching these so-called anomalous events?

The answers to these questions will synthesise an appropriate discussion on the title. By answering the research question posed above, I hope to make a convincing argument for the historical limitations of our description of the "rational, normal, expected or standard" agent in IR. It should become clear that we can, with insights gained from BIR, more correctly specify how actors make decisions, more accurately identify truly anomalous behaviour (behaviour that is not the result of systematic and consistent biases) and create more effective predictive models as well as more robust explanatory theories. This is the key contribution BIR hopes to make.

This review will attempt to piece together work that is distinct as well as show the progressive coherence of certain elements of behavioural research and the accumulation of knowledge and consensus on key issues. More tangibly, the review will use an integrative analyse=is six important works in the literature to make its arguments. Other texts are used to justify certain statements and provide evidence for arguments, but these six texts will represent the

bedrock for the argument that BIR might be able to explain previously misunderstood behaviour.

My hypothesis is that this integrative review will find that cognitive bias and cognition are key explanatory variables in grasping anomalous behaviour by actors in International Relations. By presenting IR scholars with a framework through which to process and make sense of seemingly strange actions, BIR can help to show that cognitive bias is an important piece of the IR explanatory puzzle.

However, the study might result in negative results, namely that BIR cannot explain the "anomalous" behaviour it attempts to, perhaps because cognitive bias seems not to be the core variable in these instances.

4. Research Design and Methodology

This section deals with the methods this study will employ in order to fulfil its task of illustrating the utility of a new methodological approach to the study of IR, namely the approach of BIR. This section details the methods of the literature review for the purposes of transparency, rigour and replicability.

4.1 A Note on Qualitative Methodology

Katz (2015) notes that qualitative research in the social sciences faces some important obstacles. He notes that the universal logic of "good scientific research" applies as much to empirical-observational data as it does to qualitative research. This review will attempt to fulfil the requirements of the "4 Rs", and thus present itself as a contribution to the IR literature, one that is clear, detailed and open to critique.

This review will aim to be i) representative, ii) unreactive, iii) reliable and iv) replicable**.** In order to be representative, the review will analyse – according to the criteria I will outline in the next section – literature in the field of BIR.

This literature may be said to be representative of the literature in the field of BIR to the extent that it fulfils these criteria: i) relevance (number of citations) ii) content (indicated by the phenomena observed and the methodology used) and iii) identity (self-identification of a piece of literature within a certain field – for example, an article may be published in an IR

or psychology journal). This review aims to collect literature from the field that gives a fair representation of the field on the whole.

The review collects data from sources that have already been written. As such, reactivity is not a key concern. The literature is already recorded and will not change based on my behaviour or perceptions. However, it is important to note that my subjective interpretation and understanding of the literature will affect the review. To this end, it will be my goal to accept critical feedback from qualified parties, attempt to remain objective and present the literature in an honest and open way in order to avoid unacceptable levels of bias.

Reliability concerns the information that I choose to include or exclude from my analysis. I will attempt to provide as much relevant data as possible and to give space to contrary views in order to show my relative objectivity. I will use purposive sampling based on a clear method in order to guide my selection and mitigate personal biases and convenience sampling.

The reliability and replicability of this review will be closely linked. The reliability of this review will largely depend on the accuracy with which I present the literature. I will endeavour to make my search process exceedingly clear so that anyone conducting an analysis on the literature of BIR, using my (explicitly stated) search terms will find the same literature. Consequently, it would be very simple to replicate the process of finding the literature I include in this review.

The unique element of the review consists in i) the wide range of synthesised texts in my core arguments as well as ii) my subjective analysis of the data. To my knowledge, no study to date collects, organises and discusses this range of literature in the field of behavioural IR to this extent. Conducting a literature search according to the criteria I outline in my methodology will lead the researcher to the same texts I have made use of in this paper. This of course contributes to the reliability of the study, since one can easily find and read the referenced material, greatly encouraging an honest and sincere interpretation of the data and the opportunity for critique by concerned parties.

4.2 Literature review method

The six texts which will be analysed in great depth include: "*Judgement under Uncertainty: Heuristics and Biases"* by Tversky & Kahneman (1974), "*Perception and Misperception in International Politics"* by Robert Jervis (1976), "*Behavioural IR as a Subfield of International Relations"* by Mintz (2007), "*The Behavioural Revolution and International Relations"* by Hafner-Burton et al. (2013), "*The Micro-Foundations of International Relations Theory: Psychology and Behavioural Economics"* by Stein (2017) and "*Political Psychology in International Relations: Beyond the Paradigms"* by Kertzer & Tingley (2018).

Bryman (2012), building upon work by Golden-Biddle and Locke (2007), makes the argument that literature reviews based on qualitative research should develop a coherent narrative. The aim of the following study will be to create such a narrative, one that makes a case for the utility of BIR in explaining strange behaviour in IR.

This review will aim to position itself as an integrative review.

The distinguishing factors of this type of review are (Victor, 2008):

i) Searching, identifying, selecting and abstracting data from studies in order to establish a theoretical causal effect, the size and nature of this effect, its consistency (or lack thereof) and the strength of the evidence for the effect's existence.

ii) Critical quality appraisal based on the relevance of the literature to the study, the presence of peer revision, professionalism, objective analysis and sound methodology.

iii) Narrative, iterative methods to bring together this data in order to discuss the aforementioned issues and the position of the study's research question.

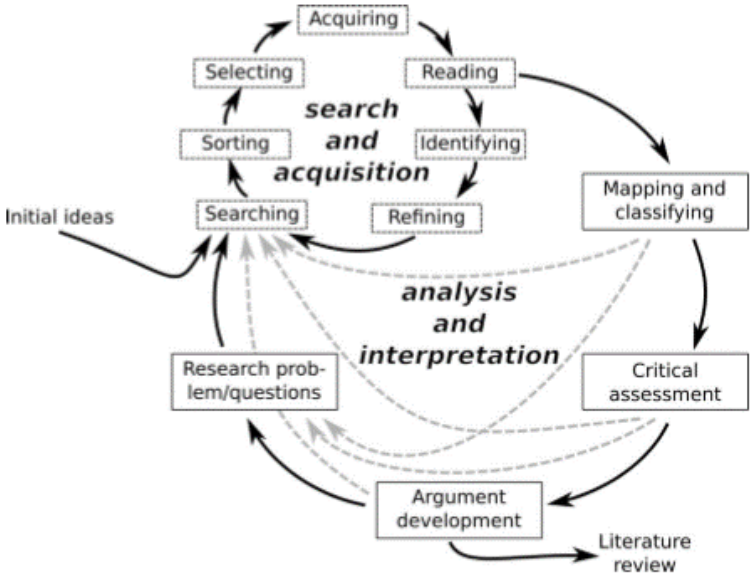
iv) A clear search strategy and explicit study selection and

v) The use of a narrative synthesis and illustrative examples to create an insightful perspective on the relevant issues.

In order to guide the process of amalgamating the relevant literature, I will use a process outlined by Boell and Cecez-Kecmanovic (2014). These authors outline a hermeneutic approach for conducting literature reviews. Below is a figure extracted from Boell and Cecez-Kecmanovic, illustrating the two hermeneutic circles that the authors describe in some detail.

This process will aid and inform the methodological process underpinning this literature review.

Figure 1: The two relevant hermeneutic circles in a literature review



Source: Boell and Cecez-Kecmanovic (2014)

In order to make the study manageable, I outlined the criteria I used to justify the selection of literature in this review. The criteria were adapted from the framework provided by Mintz (2007), which outlines what BIR consists of.

Articles included in this literature review had to meet the following criteria for them to be considered as part of the BIR literature. Though they may not have classified their studies as part of BIR, their objects of inquiry and methods fit within the field of BIR as defined by Mintz (2007).

1) The study was focused on the less-than-rational decision-making of applicable political entities (see 3 to determine which political entities are applicable).
2) The study was concerned either with the process or outcome (or both) of the decision-making process. That is, the selected study either identifies interesting information processing systems in agents or deals with the unexpected outcomes accompanied by these systems.

3) The study was centred around the judgements and choices of: i) heads of state; ii) political, economic and military elites; iii) international organisations; iv) domestic organisations; v) states; vi) political or military coalitions OR vii) the public. There exist both individual and group-level biases. At the moment, it is unclear how the two interact, but there is evidence of different systematic biases which emerge from both individuals and groups.
4) The studies must meet the requirements of proper scientific rigour. Thus, the literature that is considered relevant includes only i) publications from peer-reviewed journals, ii) publications from credible universities, institutions or organisations, including notes from meetings or conferences or iii) other miscellaneous publications vetted by credible institutions where the content has been reviewed and written in the spirit of academic professionalism, and is useful and competent.
5) Question of interest must include i) war – its initiation, escalation and termination; ii) terrorist decision-making; iii) errors in judgement; iv) peace and mediation; (v) international economic and aid decisions; vi) environmental issues; vii) negotiation; viii) deterrence or ix) democratic peace.

The search concepts guide the selection process by sieving out irrelevant literature. Articles bearing no clear resemblance to the search concepts were excluded from the literature review. The following criteria outlines the methodology used to search for and read the literature that was relevant to the review, though not all the identified studies are analysed in detail.

Table 1: Search concepts, ranked in order of importance:

Search Concept 1	Search Concept 2	Search Concept 3
International Relations	Behavioural studies	Elites
Political Sciences	Behavioural science	Executives
Social Sciences	Psychological studies	Leaders
	Cognitive studies	Nations
	Behavioural economics	Organisations
		Systems

Table 2: Comprehensive list of search terms and relevant results

Operator/s	Phrase 1	Operator/s	Phrase 2	Relevant articles identified *
	"Behavioural International Relations"			17
	"Behavioural IR"			13
allintitle:	Behavioural	AND	International Relations	8
allintitle:	Behavioural	AND	Political Science	4
allintitle:	Behavioural	AND	IR	1
allintitle:	Behaviouralism	AND	Political	8
allintitle:	Behaviouralism	AND	Politics	1
allintitle:	Behavioural	AND	Political Psychology	1
allintitle:	Behavioural	AND	Politics	6
allintitle:	Cognitive	AND	International Relations	1
allintitle:	Cognitive	AND	IR	1
allintitle:	Cognitive	AND	Political Psychology	2
allintitle:	Cognitive	AND	Politics	3
allintitle:	Political psychology	AND	International relations	4
Number of articles identified by manually searching bibliographies of important texts:				32
TOTAL:				102

*Articles searched in order of table, if the same article was identified in more than one search, it was attributed to belonging to the 1st category it was found in.

Note that all searches include both the United Kingdom (UK) and American spelling of the word: Behaviour I.e. The phrase "Behavioural International Relations" was searched using: "Behavioural International Relations" OR "Behavioral International Relations"

The literature identified by the search criteria outlined above constitute the bulk of the literature in the field of (or highly relevant to) BIR.

The six "seminal" texts I use in the review are easily identified by the above search criteria and appear in many of the searches containing these terms. They represent the most relevant of the identified texts, judged by number of citations, number of repeat appearances in the literature search, content and quality appraisal.

I list the number of texts identified in order to provide evidence for the rigour of the literature search conducted. While only a purposive sample is used in the integrative review, I have

read widely in the field and noted and been informed by the common themes and issues in the broader literature.

4.3 Ethical Considerations

This literature review is a culmination of previously published literature. It adapts, summarises and synthesises these data in order to answer its research question. This review also makes use of illustrative cases where applicable. This review does not make use of any surveys or collect information for any persons. No biological testing is required for this review. Additionally, all references are cited explicitly both in the text of the review where applicable and in full at the bibliography at the end of the study. I do not, to the best of my knowledge, require any additional consent for the use of the data in this literature review. All of the literature used and cited is available in the public intellectual domain and/or was made available to me via the library of the University of Pretoria, South Africa. This review attempts to give an objective view into the world of Behavioural International Relations and associated phenomena. My arguments, where no citation is provided, are my own, and are justified by the data provided. I do not foresee any ethical dilemmas pertaining to this work.

5. Behavioural IR: A Brief Note on its Genealogy and Evolution

Scholars of International Relations (IR) have drawn on behavioural economics and cognitive psychology to explain anomalous behaviour by political elites for over half a century. The 1950s and 1960s saw a 'great debate' in IR that emphasized the role of individual and group behaviour in international politics, as opposed to the role of institutions and political philosophy – the discussion of the first behavioural revolution. The second, beginning with the work of Kahneman and Tversky (1974), continues to this day. In 2005, a panel at the annual meeting of the International Studies Association made the case for Behavioural IR as a subfield of International Relations. This subfield is still underdeveloped, but – through its emphasis on cognition and cognitive bias as a key variable – promises to increase the explanatory power of IR and provide it with more realistic psychological micro-foundations.

5.1 The Behavioural Revolution(s) in Political Science

Before we discuss the "behavioural revolution" in IR, we need to clarify what we mean by the term 'behavioural'. Firstly, it is necessary to draw an important distinction. Behaviouralism is sometimes conflated with Behaviourism in the political science literature, and this conflation is the cause of no small amount of confusion. Behaviourism was a term borrowed from Psychology, where it denotes an emphasis on objective, quantifiable behaviour as the key variable of analysis. By analysing external agent behaviour instead of subjective feeling or experience (those things that take place inside the mind), one can – it is argued – maintain the high level of objectivity, validity and extrapolative ability required by formal science. Psychological behaviourism aimed at avoiding subjective experience in favour of experimental analyses, often with the application of a given stimuli and the monitoring of responses. This strict method has given way in the cognitive sciences to a more nuanced approach. Cognition is, and has been for some time, a term referring to the information processing systems in the mind, which may or may not be theoretically separable from the sensory and emotive systems. This development of cognition as a concept has encouraged psychologists to delve back into the inner workings as well as the exterior manifestations of the human mind. The work of psychologists in this area is suited to a fruitful dialogue with economics and politics, where we see the effects of subrational thinking affect the world, often on an international scale. The most important conceptual link between psychological Behaviourism and the political Behaviouralism I refer to is the agreement on the importance of scientific validity and the analysis of individual agents, and there is very limited connection beyond this (Simon, 1985).

In contrast to Behaviourism, Behaviouralism is an analytical and methodological approach that was formulated in the political sciences to analyse and model the actions of voters, political leaders and other relevant individuals. This approach focuses on a more scientific approach to political analysis than other theoretical paradigms, since behaviour is (sometimes) quantifiable and represents numerous data points for meaningful analysis. Easton (1962) notes that it consists of some major doctrines. Very briefly they are:

i) There exists certain observable (and predictable) political behaviours. These behaviours can be expressed accurately in theory, which should be deduced from the observed behaviour for this theory to be considered valid.

ii) These theoretical statements can be tested against observed behaviour, and therefore verified or falsified with good justification.

iii) There exist a set of academically acceptable (and unacceptable) methodological means for recording, acquiring and interpreting data. Techniques for analysis should be scientific and systematic, and should incorporate methods from the other sciences where applicable. Anecdotal evidence should be limited, for example. Contrastingly, assertions based on statically prepared and analytically assessed data should inform scientific inquiry.

iv) A certain level of objectivity should be maintained, and subjective interpretations should be clearly labelled as such. Normative judgments of values and norms should be avoided where possible.

v) There should be a close, tangible relationship between research, theory and practical problem-solving in the real world. This relationship will lay the foundation for the creation of a set of general principles, a theory of 'behaviouralism'.

vi) The political sciences should integrate with and permeate into other social sciences.

The 1950s and 1960s represent the birth of political science as a modern social science, and behaviouralism was foundational in this event (Berkenpas, 2016). Wogu (2013) notes that Behaviouralism was developed primarily through American political authors including David B. Truman, Robert Dahl, Evron M. Kirkpatrick, Charles E. Merriam and David Easton. These authors reiterate that the aim of behaviouralism in IR is to provide a view that simultaneously incorporates different theoretical approaches while creating a new lens for its analysis. This new analysis, according to Easton (1962), provides the opportunity to build a general political theory that is – as opposed to traditional historicist theory – able to formulate and test hypotheses in the here and now.

Robert Dahl (1961:772) summarises his hope for the behavioural approach in political science. He asserts that "unless the study of politics generates and is guided by broad, bold, even if highly vulnerable general theories, it is headed for the ultimate disaster of triviality."

For Dahl, the injection of a spirit of empirical inquiry had, and has today, the potential to permanently alter the study of polities, and ensure its continuing relevance for scholars and policymakers. Dahl predicts that the behavioural mood would disappear as a distinct form of – as it was seen in the 1960s – protest. Rather, he argues, it will dissipate because it will become commonplace, adopted into the discipline as an important part of the identity of political science. Dahl's vision has not been fully realised, but his hope is one I share, and my work here hopes to contribute in some small way to this realisation (Dahl, 1961).

Behaviouralism, after the 1960s, was significantly critiqued (Berkenpas, 2016). It was criticised for lacking theoretical, historical and philosophical foundations and contributions, being instead focused on experiments and simple hypothesis testing. Authors criticised that the largely positivist approach of Behaviouralism opened it up to the same limitations as positivism. These claims were deflected, because, as will become increasingly clear, Behaviouralism is distinct from positivism, though it adopts some of its (especially methodological) approaches. Anti-behaviouralists also made the claim that behavioural approaches tend towards aimless empiricism, because of their scientific mindset. Another issue, related to the above, is that, since the behavioural approach requires some level of empirical focus, it is concerned mainly with easily observable phenomena (Wogu, 2013). This is an important claim, and one that will be addressed in the study later. It is a claim advocates of BIR should be aware of, because if studies focus on convenience samples or easily measurable phenomena at the expense of more complex processes, there remains the risk of BIR being marginalised and becoming irrelevant for the most important international developments. It should also be noted that some critics were actually providing counterpoints against psychological Behaviourism and not actually Behaviouralism as it was practiced by scholars of International Relations.

5.2 The Behavioural Revolution: Diffusion into International Relations

The great complexity of the human choice environment, the inability of people to possess full information and the role of values and individual (therefore variable) experience are

often overlooked in developing 'rationalistic' models of probable outcomes, making them poor predictors of behaviour (Tversky & Kahneman, 1974). People, both political elites and the public, routinely make choices that are sub-optimal economically or politically speaking. Instead of being reliably rational, people are more accurately described as reliably subrational. The term subrational here is an important one. I, noting the work of Kahneman, prefer the term 'subrational' in favour of the term 'irrational'. Kahneman does not use the term subrational per se but highlights the importance of avoiding conflation between irrationality and subrationality.

The term rational denotes a specific technical meaning, namely that agents optimise in a given model and select from a set of choices in order to maximise their utility – and that these preferences are clear and consistent. The term irrational brings forth the image of the human decision maker as a rabid, frenetic agent. I do not wish to imply that this is the case. We are at once more than and less than rational, complicated and interesting. The term subrational more accurately denotes the truth – that the scope conditions for actor rationality are narrower than traditional models imply, which is not the same as saying actors are irrational.

Social scientists of the early 20th century largely considered human agents near-perfect in their rational decision-making. Models of human behaviour assumed we possess unbounded rationality. More recently however, these models have been questioned and these assumptions have been widely dismissed. The concept of bounded rationality now permeates new theories of human behaviour. Research has expanded into finance, game theory, international political economy, political science, behavioural international relations and other fields (Stein, 2017). Bounded rationality narrows the scope conditions for rationality, in systematic ways. The discovery that deviations from rational choice models is not all due to noise (randomness) but is in fact, in significant part, the result of bias (systematic error).

Where Kahneman and Tversky sparked the bonfire of the behavioural revolution, Robert Jervis – with his book *Perception and Misperception in International Politics* (1976) – carried that flame into International Relations. Jervis now-seminal text was one of the first of a very small group of authors to ever meaningfully – and successfully – open dialogue between IR

and psychological theorists. The connection between foreign policy analysis and the study of cognition was novel at the time of his writing. Jervis' main argument is one that I share: Political actors perceive the world in ways that differ from objective reality, and these perceptions significantly affect the international world. I go one step further, asserting that political actors hold beliefs that differ from the objective, external world not only due to perception, but also because of their reception of information. Additionally, many (though not all) of these divergences are predictable. Jervis' work will be analysed more thoroughly in my review section, but it is important here to note that Jervis' analysis was, and is, one of the "best critical syntheses of the literature available" (Eldridge, 1977)

In a foundational work for the field of BIR, Alex Mintz (2007) positions "Behavioral IR as a Subfield of International Relations". In a round-table in 2006, at the annual meeting of the International Studies Association, scholars discussed the core ideas for a new subfield in IR. Mintz notes the deviations in observed behaviour from "traditional analytic, rational, expected utility model of choice", including biases like framing effects, heuristic bias, emotional effects on decision-making and many others. Mintz also begins defining the subfield by listing six characteristics of BIR, its relevant actors, concepts, levels of analysis, methods and questions of interest. The following analysis will make extensive use of these concepts to define what constitutes BIR and inform the entire paper. Mintz (2007: 162) makes some important insights into what BIR can offer the field, and notes that "Behavioral IR is concerned with how cognitive limitations, psychological factors, and susceptibility to biases affect IR…and therefore…enrich understanding of international politics."

Essentially, the approach that BIR takes is multi-disciplinary. It uses and is informed by theoretical insights and may, at times, concern itself with the purely theoretical. However, BIR also uses empirical (usually psychologically-based) research to inform its worldview. By nature, the behavioural approach seeks to create bridges between disciplines to in order to create insights which are both verifiable and complex (Mintz, 2007).

Hafner-Burton et al. (2012) trace the more recent rebirth of the behavioural revolution within IR. It is relatively widely accepted that work by Tversky & Kahneman (1974) got the behavioural ball re-rolling in earnest in their respective disciplines – namely economics and psychology. The rebirth of the behavioural approach in IR has, arguably, piggybacked onto

these fields. Interestingly, under Kahneman, there is growing consensus among a relatively simple, 2-part model of cognition, that permeates many theoretical approaches in behavioural studies across the social sciences. This model uses a metaphor to break our thinking into two parts, fast and slow thinking, where the former is an intuitive and unconscious cognitive process and the latter is characterised by more deliberate and focused thinking. This will be elaborated upon later, when we discuss the dual concepts of cognition and cognitive bias. Importantly, there are also new insights from neurological sciences, which ground behavioural studies in observable, replicable phenomena. Hafner-Burton, et al. (2013) note that political scientists have been slow in developing the model of the subrational agent. They simultaneously illustrate ways in which research in this sphere may be moved forward and provide compelling reason to further investigate the specific attributes of individual decision-makers – one avenue this paper will explore.

6. Conclusion

The chapter began by outlining the structure of the integrative review and analysis of (mis)behaviour and bias which will take place. It identified the research theme, the research question, the hypothesis and the goals of the review. It outlined the methodology used in the review.

The latter half of the chapter gave a very brief account of the roots of behavioural social sciences and its diffusion into International Relations. It discusses the methodological and theoretical foundations of Behaviouralism, and how these foundations inform the subfield of Behavioural International Relations. The chapter notes that the subfield of BIR is still underdeveloped, but that influential authors are beginning to build consensus on how to frame and tackle the issues of cognition and cognitive bias in the international arena. Hopefully, the remainder of my analysis will further elaborate on the potential for the utility of the subfield through the building of a useful framework and analysis of the relevant literature.

Chapter 2

Cognitive Bias in IR: A Conceptual and Theoretical Framework

1. Introduction

Human beings are stubbornly complicated. On many levels, we seem to defy simple explanations. The complexity of the human decision-making process is one that influences every facet of human existence. This process influences those facets to an even greater extent when the decisions that need to be made are international in scale. Classical theories of economics and international relations endeavoured (nobly, though optimistically) to bring the human decision maker into the light of scientific inquiry, and to analyse this behaviour and identify its nature. Human beings, as in all scientific inquiry, were the subject of scrutiny, of experimentation. However, theoretical parsimony and human beings seem to be mortal enemies. We do not formulate utility functions as classical economists have us believe, nor do we base our decisions on unidimensional questions of power or even co-operation. Instead, we are nuanced, and this nuance is, by all accounts, a nuisance.

Further, the complexity of the consumer choice environment, the inability of consumer to possess full information and the role of values and individual (therefore variable) experience are oftentimes overlooked in models of decision-making in the social sciences, and especially IR. Despite the empirically supported trend of unpredictable human behaviour, only a few decades ago, fields like behavioural economics were non-existent. Camerer et al. (2004) point out that the idea of introducing insights from the psychological sciences was repugnant to most positivistic social scientists in the 1960s.

One aim of this analysis is to uncover the cognitive biases and psychological phenomena that influence our thinking beyond a cost-benefit type analysis. By incorporating understanding from different fields of study, including psychology, economics, political science and evolutionary biology, this analysis will aim to create a useful map of the key drivers of cognitive bias. The following section will note the limits of our rationality and some of the concepts that inform a position on the nature of our decision-making process.

This chapter will then attempt to create a theoretical and conceptual framework that specifies some of the conditions for rational and irrational behaviour, the scope of these conditions and potential ways to mitigate such biases. This framework is important because it will describe what I mean when I refer to cognitive bias and thus inform the integrative review of BIR that follows. Additionally, it will be a useful tool for understanding when and how cognitive biases play themselves out and to what extent these biases effect the international system.

2. Discussing Key Terms

Definitions matter. Without clear definitions, debate is impossible – or worse, incendiary – and the pursuit of knowledge is hampered. This section will use the existing literature to note the different and sometimes unclear meanings of these terms, and then define their meaning in the context of this particular study. It will also briefly note some of the qualities, contentions and characteristics of the terms. The aim of this process is to create a basis for mutual understanding.

Anomalous behaviour

Anomalous behaviour refers to decisions made in International Relations that are judged, by IR scholars, to be poorly explained by existing IR theories. The IR literature may avoid examining anomalous behaviour because it lacks the theoretical and methodological tools to do so (Stein, 2017). Examples of "anomalous behaviour" include the U.S. invasion of Iraq at the start of the century, the tendency for groups to reinforce poor decisions and ideological responses to uncertainty. What makes these kinds of behaviour anomalous is their lack of strategic rationality. That is to say, if one sets up a strategic game of a game-theoretic nature, one would not be able to predict these kinds of behaviour with classical models of choice.

Anomalous behaviour is the result of influences intrinsic in the human decision-making process, and are caused by the kinds of characteristics Tversky and Kahneman (1974, 1979)

identified in their research. These include heuristics that we use to make decisions quickly but sometimes lead to suboptimal outcomes. For example, we value the present more than the future, and would rather eat one marshmallow now than 3 in a week's time, even though 3 marshmallows would be better than 1. Of course, when we understand human behaviour, we will not need to label any kind of behaviour "anomalous". It is a term sometimes used to hide the fact that our current understanding is flawed. It is easier to say, "that behaviour is strange" rather than "my understanding of human behaviour is inadequate". It is my hope that behavioural work will allow us to call behaviour we now call anomalous merely "behaviour".

Behaviouralism

Behaviouralism is an approach to studying the social sciences that is characterised by the scientific method. It focuses on data collection, modelling and statistical analysis to observe and make statements and predictions. It seeks to examine the behaviour of individuals and groups rather than the nature of institutions in order to explain their behaviour as it relates to the domestic and international system (Easton, 1962 & Wogu, 2013). Behaviouralism draws its theory not from deductive or axiomatic sets of statements. It does not intend to present a complete and entirely unified theory, at least not in its early stages. It seeks merely to investigate the way in which the human decision-maker comes to his/her decisions.

Behaviouralists, like constructivists, look beyond the rational. Behaviouralism is nonrational, or quasirational, in the sesne that it looks to "prevailing ideas, norms, heuristics, and logics of appropriateness as determinants of individual and social choice processes"(Hafner-Burton et al., 2017). Historically, Behaviouralists were skeptical about the historical, philosophical approaches of political scientists. They shared the belief that new methods existed or could be created that were systematic and testable, and would improve the study of politics. Some commentators indicated that Behaviouralism was nothing more than this belief, a statement without a theoretical and methodological framework (Dahl, 1961). Behaviouralism certainly evolved past that limited point, but has never quite found for itself the same narrative thread that Realism, Liberalism or Marxism enjoy. Behaviouralism never can, because it does not deal with a coherent story of the human race, but the human race itself, which is far from coherent, simple or static.

Behavioural International Relations (BIR)

BIR is a subfield of IR that focuses on phenomena where behaviour deviates from the standard rational choice model. It is centrally concerned with how cognitive limits, psychology and human susceptibility to bias affects international politics. Its methods focus on experimentation, computational modelling and statistical analysis, but some qualitative analysis also finds a place in the field, so it is not merely a reversion to behaviouralism, but rather an extension of it (Mintz, 2007 & Hafner-Burton, et al., 2017).

Behavioural IR has existed since the 1960s with its members publishing in outlets such as the Journal of Conflict Resolution, International Interactions, Journal of Peace Research, Conflict Management and Peace Science, and International Studies Quarterly" (Mintz, 2007:166). What is new, however, is the opportunity to study political elites at a more precise level and achieve more robust results. BIR may become more relevant than ever before, as "the intervening causal mechanisms within states are subjected to closer scrutiny in the analysis of relations between countries and other organizations in world politics" (Mintz, 2007:169). The field is, as yet, underdeveloped, but the rationale for its utility is strong, and it represents an opportunity for scholars to increase their collective understanding of the most complicated subject that we know of: the human being.

Beliefs

Beliefs, in the context of this study, refer to the **"**Individual, persistently divergent, perceptions of the world we jointly observe" (Bénabou, 2015). Our beliefs are often distorted, and affect our decision-making in ways that harm our economic and political systems. Examples of distorted beliefs include overconfidence or groupthink.

Hafner-Burton et al. (2017:1) note the importance of beliefs in International Relations: "The Clinton and Bush administrations did not differ substantially in their information about Iraq. But Bush administration officials—and the president himself—did hold beliefs that differed substantially from those of their predecessors, and those beliefs had profound effects." Beliefs are one lens through which we process information. In the world of international affairs, there is seldom a clear choice. Strategy in an uncertain environment requires leaders to act based on more than mere facts, to be pre-emptive we must expect, predict and pre-empt. This pre-emption requires leaders to make educated guesses about the meaning of

certain actions, and this meaning can only be garnered from beliefs. Since beliefs can be misleading, it is important to hold council with those whose beliefs differ, and with those whose experience or relationships allow them to have better informed beliefs about the meaning of the actions of an adversary. In this way, leaders can attempt to mitigate any potentially rash decisions, and attempt to see the most likely outcomes, as opposed to the outcomes that they believe are most likely – which are by no means necessarily the same.

Cognitive bias

Cognitive bias is a broad term referring to the tendency for human decision-makers to perceive, process, recall or reason in a rationally deficient, biased or incorrect way. Confirmation and memory bias as well as a long list of heuristics fall into the category of cognitive bias. This study will focus on those cognitive biases most relevant to International Relations (Kahneman & Tversky, 1974 & 2013; Farnham, 1994).

Of course, critics note the subjective nature of 'bias'. What one considers biased decision-making, another may consider reasonable and even desirable. I think critics who argue along these lines conflate ideas. Cognitive bias is not a set of values, it does not judge behaviour in the sense that unbiased behaviour is good and biased behaviour, bad. Cognitive bias defines rational behaviour in the strategic, game-theoretic sense. Where there is information, and insofar as it is possible to rank order the likelihood of certain events and thus assign probabilistic, weighted value (often monetary value), then – under these conditions – we can define biased and unbiased behaviour. Implicit in the analysis of behavioural scholars is the assumption that information provides us some measure of evidence for weighing the likelihood of an event and the costs and benefits of that event. Further, cognitive biases are the patterns that scholars have identified which illustrate the systematic and predictable ways that human beings deviate from correctly grasping the probabilistic outcome and its corresponding utility. The expected utility model of choice gives us a benchmark from which to measure 'faulty' reasoning.

These games become exponentially more complex when they are played in the international arena, and so it becomes even more important to understand how to mitigate any biases that may leave one open to undue risk, cost an inordinate amount or endanger the peace between nations.

Expected behaviour

Expected behaviour refers to the most likely outcome in an analysis given core assumptions of full information, rationality as well as transitive, clear and constant preferences and the presence of an unambiguous problem. Full information refers to a situation where all involved actors know the preferences and beliefs of other actors. This is rarely the case in the international arena, and so the economic model of rational choice and expected behaviour begins to leak from a myriad of holes. The term *expected behaviou*r is often used in the field of behavioural economics, which informs much of the literature in BIR, and is hence relevant to this study.

In the review, expected behaviour may also refer to the most plausible outcomes given a specific theoretical presupposition. E.g. In IR, an advocate of Liberalism would struggle to predict the success of Donald Trump in winning US elections in a world of (at least in most of the developed world) free speech, tolerance and increasingly politically correct values (DellaVigna, 2009 & Stein, 2017).

Heterogeneity

In behavioural economics, heterogeneity refers to the tendency for decision-makers to vary widely in preferences and beliefs. As a result, individuals respond differently to economic and political stimuli, sometimes in ways that are difficult to predict. The observation of heterogeneity is little more than the observation that human beings are individuals and vary across multiple dimensions. An important question, and one that is not yet satisfactorily answered in the literature is this: How is "individual heterogeneity aggregated into collective choice" (Stein,2017:256) decisions?

There is, however, a fine line to be drawn. Some may question the validity of saying that human beings, in general, show certain cognitive biases. Didn't we just assert that every individual is different? While individual beliefs and preferences may change substantially, there seem to be evolutionary-biological mechanisms that allow us to make certain decisions based on heuristics, and which are common to every person. In this way, these beliefs and preferences are predictable or systematic (Kahneman and Tversky, 1979). Some of these mechanisms can be overcome, others must be tolerated, and others are welcome, because they allow us to make sense and meaning in a complex world.

Heuristics

Heuristics is a term first brought to popularity by Kahneman and Tversky (1974,1979). They showed, experimentally, that there are certain rules of thumb which we use to answer questions or make assessments that are both rapid and erratic.

Heuristics can be described as cognitive shortcuts that allow us to make decisions quickly, by drawing on past experiences and anecdotal evidence. These cognitive shortcuts empower us to make tasks less complex, but may also lead to systematic errors in judgment. Heuristics, it is posited, may be an evolutionary development. We jump at a noise in the grass because it may be a snake, our brain does not weigh up the evidence to make the decision. The specific movement of the grass, the speed of the wind, the sound of the rustle, all of these become irrelevant. There are certain instances where the probability of being wrong is large but the risk is larger, and so we jump out of the way whether there is a snake in the grass or not. Such thinking serves us well at times, and at others, it leads to costly errors.

Mintz (2004b) makes the argument that heuristics are often used as part of a strategic decision. The set of possible actions is often narrowed first by heuristics, before a more rigorous analysis is conducted. An alternative to this may be the use of a simple algorithm that narrows down decisions based on a formula as opposed to the use of less uniform and strategically inefficient heuristics.

Integrative Review

An integrative review is a review type characterised by an integrative approach to creating new knowledge, a theory development agenda and the creation of sophisticated syntheses to understand phenomena. Stages include searching for, appraising, synthesising and reporting on the included literature (Bryman, 2012 & Victor, 2008).

Whittemore and Knafl (2005) present an updated methodology for the integrative review. The integrative review "summarizes past empirical or theoretical literature to provide a more comprehensive understanding of a particular phenomenon" (Whittemore and Knafl, 2005:546). Integrative reviews possess the capacity to inform policy, practice and research. Data reduction, display, comparison and conclusion drawing are all central steps to the integrative review.

It is my aim that this review fulfils these criteria by expelling the noise surrounding the behavioural approach and focusing in on the key concepts and issues. It is also my hope that by presenting a summary and discussion of the key texts that the review both displays and compares relevant work.

Preferences

I will use this term in the (primarily) economic sense. Preferences refer to the set of assumptions and corresponding ordering scheme of a particular set of options. This ordering system is based on the degree of gain, utility or happiness different alternatives afford a decision-maker (Lichtenstein & Slovic, 2006).

Preferences extend from favoured foods, preferred time-horizons (shot or long-term) to preferences for war or peace. It is difficult to understand the unspoken and associated meanings of preferences without having some understanding of the notion's use in economic theory. Economic theory ascribes "mental states, such as beliefs and preferences, to the agents in question and (ii) showing that, under the assumption that those agents are rational, the ascribed mental states lead us to predict, and thereby to make sense of, the behaviour to be explained" (Dietrich and List, 2016:1-2).

Hafner-Burton et al. (2017:6) note that "preferences are simply the actor's subjective rank-ordering of the terminal nodes or outcomes of the strategic interaction". Standard models of actor preferences certainly allow for individual variation among actors. Behavioural work in IR actually has much to offer formal rationalistic models here. Empirical evidence of different actor preferences can contribute to more specific models of actor preferences and, by extension, better understand and predict, actor behaviour. Where evidence indicates strategically unfavourable decisions (e.g. social preferences like altruism or nationalism), the formal rational models of human behaviour are fundamentally threatened.

Traditional International Relations Theory/Traditional IR Paradigms

Theories of international relations that posit the state as the central actor in IR and emphasise variables such as material gain, military power and ideological belief. They include (among some others) Realism, Institutionalism, Liberalism, Constructivism, The English School, Marxism and Feminism (Walt, 1998 & Griffiths, 2007).

The traditional IR theories have both "important strengths and serious weaknesses" (Ferguson, 2015: 3). There are areas in which their respective toolkits are effective at dealing with a given research subject. I do not wish to suggest that traditional IR theory does not enhance one's understanding by presenting a valuable theoretical perspective. However, in the case of BIR, traditional IR theory fails to adequately deal with certain events. Those events are dealt with in this study, and illustrate that BIR plugs the methodological and theoretical holes that threaten to sink the metaphorical boats of the once dominant IR paradigms.

3. A Brief Account of Cognition

Cognition is a difficult term to pin down. It is by no means self-evident what is meant by the term, since both relatively subtle and clearly overt differences in definition and emphasis exist between fields and individuals. I will aim in this section to give a brief account of the term as it applies to this review.

What is cognition, exactly? It may be described as a signifier of the complex interactions happening within the human brain. Of course, while correct, such a broad definition of cognition tells us little about the specifics of the term, especially as it applies in the sphere of International Relations. At the most basic, biological level, cognition is nothing more or less than information processing, including both the information processed by the central nervous system as well as the brain (Århem & Liljenströmb, 1997). Such a simple characterisation masks the vast and mysterious complexity of this endeavour. Cognitive psychologists and neuroscientists attempt to investigate these unmercifully complicated mental processes.

Cognition is also an approach to the study of the mind, especially the mental activities or procedures of obtaining knowledge and understanding through thought, experience, and the senses. One might think that the behavioural revolution – being so intimately involved with the concept – begun with a strong theory of cognition. This is not the case, but as Hafner-Burton et al. (2017:8) point out, "one gradually emerged out of this work and has ultimately been given a biological foundation in new brain research." Thanks largely to the work of Tversky and Kahneman, a dual-level model of cognition as emerged. This might be referred to 'fast and slow thinking' or 'hot and cold cognition'. Kertzer and Tingley (2018:8) make the

case that behavioural research is beginning to move away from cold cognition to hot cognition. This represents a shift from work on information processing to work on how emotional influence affects decision-making. Cognitive scientists have a history of using controlled experiments, and there have always been some questions as to whether or not laboratory experiments are transferable to the real world, and to what extent. Using cognitive experiments with political elites increases the validity of the applications of cognitive psychology in International Relations.

The field of cognitive psychology has been instrumental in aiding the goals of political scientists. Cognitive psychologists investigate questions of perception, learning, experience and memory. They investigate why objects appear farther away on foggy days, why we forget the names of lifelong acquaintances but easily recall embarrassing or ecstatic memories and why multinational companies spend fortunes on advertising (Sternberg, et al., 2012). The answers to these questions often provide valuable insights for the political scientists. Think of advertising, for example. The availability heuristic, described in Table 3 on page 39, explains the power of advertising. When we ask which phone to buy, and which ones will be reliable, we think of brands we can easily recall. This bias, the assumption that what we can remember is the most relevant information, leads multinational companies to spend on advertising, and political elites to advertise their accomplishments and hide their failures.

While there exists sets of theory, explanations and laws among groups of cognitive scientists – whether in biology or psychology – what characterises the study of cognition among scholars is (similarly to BIR) an "approach to the study of the mind rather than…permanent theoretical commitments" (Von Eckardt, 1995:15). To this point, the study of cognition is still young and developing. It may be described as an immature science rather than a mature one, still tentative about its theoretical assertions, and open to influence from a variety of sources. Behavioural International Relations has much to gain from cognitive scientists, and, if my hopes come true, perhaps it will repay that debt with novel contributions to the field in the future.

In the following chapter (Chapter 3), I will attempt to trace the development of cognition in International Relation. While there is, as yet, no unified theory of cognition in IR, one is being developed. As more work is done on cognition in the field of IR, the methodological

approaches of behavioural scholars, the terms that they use to construct their analyses and frameworks are becoming familiar, and may develop into something approximating a unified theory in the coming years.

4. On the Limits of Human Rationality

The legacy of the enlightenment largely permeates our identity as modern human beings. The ideals of individual rationality are central to our conceptualisations of ourselves. This outlook, however, is not a result of the enlightenment alone. Indeed, millennia ago, Aristotle classified human beings precisely as the rational animal. The idea is both powerful and, at least in part, misleading. Of course, human beings are capable of rationality, this paper wouldn't exist otherwise. But to posit the human being as a rational animal misleads, for we are not merely rational. At the very least, we do not comply with the assumptions of classical rational choice theory. When making decisions, people do not create utility functions to judge the worth of alternative choices. We do not systematically weigh choices, often we lack the time, the energy, the information or the capacity.

Kahneman (2012) distinguishes between two major modes of thinking. We might think of the human decision-making process as having two major operations. We perform most activities without conscious thought. Unconscious cognitive processes like identifying faces, recalling Rome as an Italian city, or the unthinking use of a home language all require computational power. Indeed, it is still not quite possible for a computer to recognise faces or interpret languages with the ease of the human mind. If we are to be defined by what we most often do, we are more accurately described as instantaneous thinkers than rational, deliberate ones. When such thinking fails us, as when we are required to calculate mathematical equations, or to reason out the logic of a particular puzzle, we turn to "slow" thinking, the kind we classify as rational.

Our fast-thinking, one might call it our "gut-feeling" is fallacious on many counts. The primordial mind – it is argued – places more value on survival than truth. The default human cognitive process, the one aimed at keeping us alive, is not rational, it is pragmatic. This "fast thinking" side of our minds looks only for readily available information, failing to search for

additional information. Other cognitive failures are readily identifiable. For example, we tend to weigh vivid experiences more heavily than mundane ones (regardless of objective importance), the need for coherent narratives because of our propensity for telling and recalling stories and our tendency to be compelled by natural intuition, even when subjective feeling runs counter to evidence contrary to our current beliefs. For these reasons and more, human beings are not very good at understanding probability intuitively. We consistently over-estimate the chances of unlikely but interesting or horrific events and under-estimate the chances of failing in a new business venture. We dismiss statistical randomness. Casinos display roulette counters that show the sequence of previous numbers, as though this could affect the future ones.

It seems that the role of beliefs in decision-making is pivotal. In many ways, rationality is subjective. Rationality, acting in one's own interest, is greatly affected by one's beliefs about what is best for oneself and for others. These concerns are linked closely with subjects of religion, culture and value. Sears and Funk (1990) illustrate that surveys on public opinions, social experiments and ethnographies show that individual preferences are significantly determined by altruistic desire, ideals of justice and fair play and desires for expression. Inglehart (1990) also asserts that similar studies show that preferences changes across time and vary between individuals, contexts and populations. These preferences change so significantly that what is deemed rational for some seems inconceivable to others. Relying on a model that presumes to predict the behaviour of everyone is thus highly inconsistent. Importantly, conventional theory regarding so-called rational choice asserts that preferences are derived from observable evidence. Beliefs, in this case, are assumed to be the product of rational deductions about that nature of perceived reality. Empirical evidence in social psychology and other fields discredit this belief. Many individuals, if not all, are shown by empirical literature to show selective bias about information. Beliefs can often be contrary to observable evidence, and a host of emotional and heuristic shortcuts to make decisions demonstrate no small amount of bias (Rabin, 1998).

Game theory research indicates that strategic interaction is often overlooked in traditional decision models and provides room for behaviour that is not wholly self-interested. Chai (2000) points out that "many types of structural assumptions will generate a large set of multiple equilibria or no equilibrium at all", and that there is an ever-growing literature to

suggest that conventional behavioural assumptions (rational choice models) predict behaviour incorrectly.

Behavioural scientists have identified hundreds of cognitive biases. It is not my task here to identify and explain them all. It is enough for this study to note the impotency of pure rational choice models and the assumption that our thinking is systematic, reasonable and scientific. It can be, yet in many ways – and often without the awareness of the decision-maker – our thinking is erroneous, subjective and relatively unconcerned with data.

5. Cognitive Bias in International Relations: A Theoretical Framework

Scholars have thought about how decision-making might affect International Relations in the past (Hafner-Burton, et al., 2017). However, there is a growing consensus on how to conceptualise these processes and their effects. This theoretical framework will aim to create a map to identify the existing theoretical attitudes towards the idea of cognitive bias. It will focus on those theories which are most relevant to the field of IR and appear to affect the international world most powerfully.

Understanding cognitive bias requires modifications to familiar paradigms. Of course, traditional IR theories differ, but – inasmuch as they are rationalist theories – they share certain assumptions about the agents in question and the environment around them. These assumptions are challenged by Behavioural IR. These challenges, substantiated by literature the behavioural sciences, will lay the basis for the rest of the discussion.

What does the existing behavioural literature make of cognitive bias? How is it played out in the realm of IR? And what processes and concepts do we currently use to conceptualise the phenomenon? These are the questions I will briefly address. Their answers will justify the frameworks I outline thereafter, which attempt to synthesise and simplify the core insights from previous literature.

The study of International Political Economy (IPE) has offered some interesting insights into the concept of cognitive bias. Theoretically, in an open economy environment, individual

preferences with respect to trade policies should be well-defined. For example, in cases where the factors of production are mobile, those who possess scarce assets should favour protectionism (they have the potential to gain the most financially speaking). Conversely, when factors of production are industry specific, those who are employed in a comparatively advantaged sector would be expected to favour free trade, since this competitive advantage will make their product/service highly competitive worldwide. This seems not to be the case. Individuals make decisions, not on their place in the market, but are influenced more strongly by emotive, psychological and even biological factors. In a word, individual preferences and beliefs seem to matter more in these cases than issues of individual net gain.

Behavioural scientists have, in the recent past, been focused on using empirical methods to examine preferences, beliefs and decision-making process to adapt expected utility and game theoretic models. A 2-stage model of cognition has been widely accepted among behavioural scholars. This model has been pioneered and gained traction under the scholarship of Kahneman & Tversky. This 2-stage model breaks down our decision-making processes into 2 broad categories – as discussed above. Specifically, it breaks down decision-making processes into either i) fast-thinking or ii) slow-thinking, where the first is instinctive and relatively intuitive and the second is more calculated and deliberate. This distinction is useful for understanding when and how human beings display biases and flawed heuristics. In some circumstances, for example, stage 2 slow-thinking can mitigate flawed heuristic processes but requires more time. Thus, this model helps us to distinguish different cognitive barriers connected with time, informational or stressor constraints. We are not equally biased all of the time. In other words, circumstances matter, and this model helps to show how.

Cognitive bias is a pervasive concept and it has been mapped in detail in numerous studies prior to this one (Montibeller & von Winterfeldt, 2015; Overall, 2016; Kahneman, 2012 and Mousavi, et al., 2016 provide some particularly useful conceptualisations). My aim here is not pretence. I am unlikely to provide a framework that is truly novel, or academically superior. Instead, I aim to make these concepts accessible by removing some of the technical jargon in the literature. Indeed, BIR as a method does this very often. Where psychological or economic nomenclature makes it difficult to understand phenomena, I have simplified it in order to bring the research into the domain of IR. With that in mind, Montibeller &

Winterfeldt (2015), who bring together a large amount of behavioural literature for the purposes of risk analysis, outline a taxonomy for understanding cognitive bias. This framework, first proposed by Arkes (1991) breaks cognitive bias by their psychological origin. Briefly, these errors arise because of:

i) Association-based (AB) errors: These errors are the consequence of unconscious mental associations like the tendency to recall traumatic experiences more easily than ordinary-everyday ones.

ii) Psychophysical-based (PB) errors: Driven by the erroneous mapping between physical stimuli and mental responses. For example, we make a choice depending on how the consequences are framed. We prefer a choice that is framed in terms of gains rather than losses.

iii) Strategy-based (SB) errors: Denotes instances where decision-makers use strategically sub-optimal strategies, e.g. where there is an analytically clear 'best choice', the agent makes a different decision. These are the most easily identifiable and solvable type of biases. For example, a bad chess or tic-tac-toe move, where the best move is identifiable and verifiable with the correct understanding of the rules and possibilities of the game.

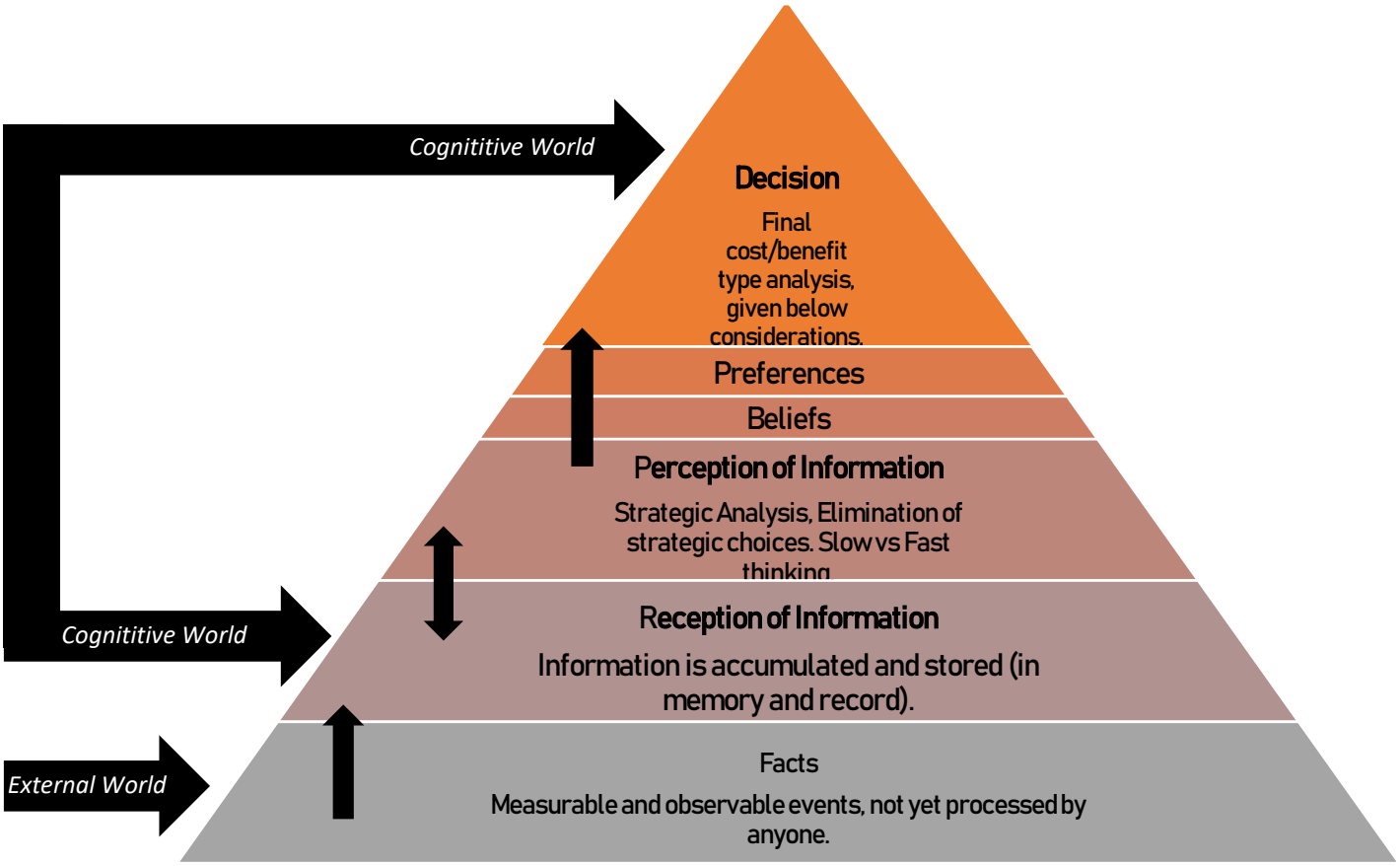
I will make use of these concepts to inform my own conceptual framework but will attempt to simplify the process so that we may identify the biases both by type and by their order in the cognitive process.

Other behavioural theory proposes decision-making frameworks which focus on the role of beliefs and attitudes. This theory of reasoned action (TRA) and the related theory of planned behaviour (TPB) have been used extensively in decision-making studies (Sheppard, et al., 1988 & Montano & Kasprzyk, 2015). Behaviour, in these models, is the culmination of cognition, judgment, attitude and intention. This step-by-step model is useful for showing that attitudes and beliefs (more than judgement) are good predictors of action. The model, however, struggles to explain the actions of decision-makers that do not correlate with rational choice. This arises where agents, given their stated preferences, beliefs and judgements act incongruently with those factors. The model fails to account for biases and errors, and those instances where behaviour runs contrary to predictions are not sufficiently

explained (Overall, 2016). For this reason, it is not sufficient merely to adapt these models for our work here in IR. Rather, in order to analyse specific systematic errors in judgement, it is necessary to adopt some insights from other behavioural studies but to adapt them with these errors in mind.

6. A Conceptual Framework for Cognitive Bias and Existing Applications

Figure 2: A Conceptual Framework for the Notion of Cognitive Bias



Source: Authors own work, adapted from a range of behavioural systems, noted above.

I will use the basic framework above to interpret and explain cognitive bias. It is constructed as follows. First, information is obtained. I will call this process the reception of information. This process begins with the presentation of information to a decision-maker. It does not involve anything more than the presence (or lack thereof) of a piece of information. Consider a fictional scenario. The trusted advisor of the president runs into the room and announces,

“Sir, (insert country name) have mobilised their troops, they’re mobilising on the border!” No decision has been made, indeed the information hasn’t really been considered yet, but the cognitive process has already begun. In all likelihood, whatever the evidence from the facts, the president in our example will assume the amassing army is a large threat because of the reaction of his advisor. Thus, “raw data” is never really “raw”. The source, manner, point in time and vividness of the information all play a role in how the information is later processed. Examples of the way the reception of information affect our cognitive process are: framing effects, negativity bias, the availability/existence of information and recurrence of information.

After obtaining or being presented with information, we process this information. This step I will refer to as the perception of information. We undergo either a stage 1, fast-thinking process or a stage 2, deliberative process. Which process we undergo depends on the complexity of the problem we are presented with or on the time-frame within which we must act. For example, shorter timeframes necessitate that we use fast, intuitive processes while some decisions can be made after deliberate thought. Continuing with our previous example, the perception of information is more likely here to be a 2-stage deliberative process, whereby other information is analysed in order to try and identify the best course of action.

The perception of information stage of our cognitive process is prone to a wide range of heuristics and biases. In the realm of IR, some of the most relevant of these biases and heuristics include: The Poliheuristic Bias, Choice Paralysis Bias, Emotional Bias, Wishful Thinking Bias and Faulty heuristics in strategic problems. These are biases inasmuch as they alter outcomes away from what is expected by rationalistic models (based essentially on possible net gains). See the Table 3 on page 39 for more detailed explanations of these concepts. Scholars including Daniel Kahneman are pessimistic about the possibility of correcting individual bias and the individual level. Establishing set behaviour patterns, creating diverse and transparent environments and establishing protocols are all ways in which we may mitigate individual bias and a more environmental and systematic level. Though these possibilities are apparent, there is surprisingly little evidence that organisations are aware of the need for such system-wide changes in order to mitigate (to a limited extent) our individual biases (Kahneman, 2012:254-264).

Once the data has been processed, we act – all things considered – and choose the most correct (in the case of, say, puzzles or mathematical problems) or beneficial (strategic problems) manner. This step is the decision step of cognitive function. It is the step that takes the processed information, which has been sieved and analysed. It is perhaps not truly distinct from the perception phase, but we will imagine that it is for the sake of this framework. The decision stage, in my framework, will include biases caused the behavioural terms "perceptions" and "beliefs". These terms often influence the decision-making process subconsciously and despite rational analysis. They might be said to be "innate" in nature, they make up part of the value system or ethic of the decision-maker.

Preferences are the subject of much debate in rationalist and behavioural models. Actors appear to assess risk, discount the future and assess social preferences differently than rationalist assumptions predict. Agents tend to weight losses more heavily than equal gains, value the future less than the present and are more inclined to share and trust than rational choice theory predicts. Preferences appear – in the behavioural literature – to be structured in predictable ways, though of course there are exceptions. Preferences, as I use the term in my conceptual framework refers to these types of phenomena. They refer to the manner in which certain elements of a decision are weighed.

Beliefs, in the behavioural literature, refer not only to the ideas that individuals have about the state of world events, but also about the ideas that individuals have about the beliefs of other players in the game. These ideas are rarely clear, or entirely correct. After all, our own minds often trick us, the mind of the other can utterly baffle. Beliefs, in our definition, includes the pervasive problem of misperception. Beliefs are shaped by available information, even if that information is inappropriate, incomplete or factually incorrect. Strong ideology, which drives single causation models may cause experts to overestimate threats because of their ability to easily imagine the causal connections that may lead to conflict. Beliefs in regard to oneself are also vital in the work of behavioural scientists. Overconfidence is an established problem in IR, and beliefs about one's own capabilities may be the cause of deadly conflicts.

I have attempted to construct a simple and accessible illustration of the cognitive bias in information processing. Facts (think about a video camera capturing an object or event)

present themselves to us. We receive this information, and we recall it in biased ways, whether in the amount or type of information that we recall. Once we have stored this base information, we analyse it, again – in predictably biased ways. We strategically analyse our dataset, eliminate certain strategies based on broad perceptions of possibilities. There exists an interplay between the reception and perception of information, where perception may actually affect what and how an event is remembered (i.e. affect the reception stage). Thus, information may be modified after the fact, based on psychological biases, and are then factored (altered) into our perception. Finally, preferences and beliefs alter our final strategic analysis and decision to take a certain action. This represents the most basic explanation of the conceptual framework that I've constructed. However, there exists a slightly more interesting point of view.

The external world delivers us events. Bringing these events into our cognitive structure immediately enriches it. Cynics might say we pollute it, but it is more promising (not to mention fascinating) to view our cognitive action on the external world as an enrichment. One that is, no doubt, prone to errors so far as objective judgements or perfect computations are concerned, but which can also intuit wisdom which is, as yet, perhaps not codified in scientific literature. Heuristics (cognitive shortcuts that help us make decisions quickly) are often effective, and, evolutionarily speaking, practical. For example, loss aversion and the tendency to remember trauma probably has played an important role in the continued survival of the human race. When stakes are high, it is prudent (for survival) to avoid risks in the face of uncertainty. The fact that, nowadays, risks are rarely fatal, does not really register with the human mind, and so we must contend with the issue and attempt – often – to overcome it. Of course, this ontological and phenomenological issue is not new. Exactly how much access we have to scientific reality is contested. I argue, in my framework (implicitly), that we have very limited access – at least intuitively – to the objective external world. As soon as we perceive events, we interpret and affect it. This, of course, is the postmodernist claim. I vary however, from the postmodern perspective, in an important sense. My conceptual framework does not subscribe to the idea that human beings interpret things in an infinite number of ways. There exist an infinite number of interpretations – this is true – but the human mind, pragmatic as it is, values those interpretations of reality which make

sense to it. These may vary slightly from objective reality, made available to us through scientific inquiry, and they constitute our predictable bias.

Of course, it is important to note that in many respects these processes all occur simultaneously, in a more nuanced way or in a different order than I have listed them here. For purposes of creating a map of the process and the phenomenon, however, framing the issue in these three steps is useful. The results of this somewhat complex cognitive process are dependent on the situation at hand. They can be dramatic and dangerous – for example, when overconfidence in a state leader begins a conflict, or they can be relatively benign – for example, if a leader's bias causes them to say something stereotypical or politically incorrect. The table below is a summary of the main biases and heuristics that seem to feature in IR.

Table 3: Main Concepts and Applications of Cognitive Bias in IR

Stage of cognition & Related Biases	Explanation	Applications in IR
Reception of Information		
Framing Effects	Tendency to judge information on how it's presented rather than remember the facts.	Leaders frame other states to legitimatise policies. E.g. To justify occupation of a non-democratic country. Citizens thus perceive the occupation as legitimate.
Availability & Repetition of Information	Often repeated and/or easily available information is seen as more relevant, useful or true than less available data.	Leaders act with incomplete\incorrect information. E.g. US misperception of counter insurgency strength in Iraq.
Negative Emotion Bias	Anger and/or fear reduces the amount of data collected and analysed before deciding.	Fear may have influenced voters to vote Trump in 2017 U.S. elections.
Understanding Bias	Humans remember what they understand more easily. Thus, information which is pertinent, but complex or confusing is dismissed in favour of more understandable data when collecting information.	Simple causation models are preferred by political experts.
Confirmation Bias	The tendency to collect, interpret or remember information that agrees with and confirms one's current beliefs, which leads to errors in analytical thinking.	Leaders are likely to collect and cite news that is coherent with their belief systems and congruent with their line of action.

Information Avoidance	"A growing theoretical and experimental literature suggests that…[there is] an incentive to avoid information, even when it is useful, free, and independent of strategic considerations" (Golman et al., 2017:96).	Avoiding information, especially noise, can be strategically beneficial. However, sometimes it is avoided simply because an agent might consider the information to carry negative utility. An example may be a wife who does not wish to know of a husband's affair, or a politician who does not wish to uncover the corruption of a colleague.

Perception of Information

Poliheuristic Bias	Leaders employ a range of heuristics to narrow large sets of choices. The rejected choices fail on at least one critical dimension. Leaders then decide from the remaining set of choices.	The democratic peace holds, in part, because it is politically costly to go to war with other democracies, but not undemocratic states.
Faulty Strategic Heuristics	There are constraints on the amount of strategic thinking agents do. Thus, equilibrium is not always found in strategic games.	Experienced political elite may play strategic games better. Political elites need to beware of reasoning by wrong analogy, like viewing Iraq through the same lens as Vietnam.
Choice Paralysis	Large and complex sets of choices lead to worse strategic decisions.	Political decision-makers may be hesitant to deviate significantly from the status quo because of complex sets of choices. For example, the lack of effective intervention in the Rwandan Genocide.
Anchoring and Adjustment Heuristic	Estimation of an outcome or value is based on an initial value or piece of evidence (an anchor) and not well-adjusted to predict the correct value or outcome. This anchor is unjustifiably assumed to predict a particular outcome.	Political Experts may be prone to using certain evidence to infer likely outcomes. This may be why political experts are poor predictors of future political outcomes.
Emotional Bias	We recall information congruent with our current emotional state more easily than other information. We also more positively assess the odds of a certain endeavor when we're in a good mood, and judge situations and other individuals in a more positive way when we are happy. The opposite is also true. Emotions also tend to exacerbate strong beliefs and can reinforce them, even if these beliefs are ill-founded.	Escalation of conflict and beliefs of the enemy is prone to emotional bias. Extreme ideological response to uncertainty and conflict likely related to emotional state. For example, the comparatively extreme response to the tragedy of 9/11 may have been driven by emotions of fear, anger and revenge as much or more than strategic reasons of security and power.

Decision

Hyperbolic Time Discounting	The tendency for individuals to discount the future at a growing rate. Thus, issues in the far future are considered disproportionately little in present decision-making.	Applications for analyzing short-term democratic systems versus long-term autocratic ones. Wide applications to issues of climate change and international co-operation for mitigating future crises. Hyperbolic discounting implies that climate change is probably considered less important by us today than it should be, given that it may constitute a global crisis in the future.
Ambiguity Aversion	Strategic decisions where probabilities of failure and success are unknown often present high possible payouts. Individuals often fail to choose ambiguous strategies despite the potential for better payoffs in the medium to long term.	Politicians, nationally and internationally, do not vary policy enough in the face of uncertainty. Thus, they may choose familiar policies even if they are suboptimal.
Inequality Aversion	Subjects show a tendency to deviate from pure selfishness in experimental settings. There is a tendency for many people to value equality (or fairness) more than efficiency.	Explains phenomenon of foreign aid and other altruistic state actions.
Overconfidence & Wishful Thinking Bias	The tendency for individuals (especially experts) to overestimate their ability to predict future outcomes. Also refers to the tendency for individuals to overestimate the chances of success in a particular endeavor, even if they know the statistical data. The wishful thinking bias refers to the tendency for induvial to rate the probability off an event more highly if they wish it to occur. This contributes to the overconfidence bias.	Political experts tend to forecast more poorly than we might assume. Further, these individuals believe their forecasts will be much more accurate than they are when examined.
Endowment Effect, Loss Aversion and Status Quo Bias	People value things more simply because they own them. This applies both to material goods but also, to a more limited extent, to ideas and beliefs. People also weight losses more heavily then equal gains, a corollary is that they prefer the status quo because potential losses loom larger than equal gains to be had from changing the status quo.	Leaders may be hesitant to bring about change because the risk of loss weights more heavily on their minds than the potential gains. E.g. the economic losses of moving away from fossil fuels seem terrible even though there are great economic gains to be had from the renewable energy sector.
Sunk-Cost bias	Individuals tend to continue a particular course of action because they have invested previous time and resources in achieving a particular goal – even when current circumstances make it unfeasible to do so.	Leaders may take too long to end a conflict abroad despite the negative sentiment of the population and high costs, e.g. the US-Vietnam war.

Adapted and Integrated from the following sources: Kahneman & Tversky (1979 & 1984), Kahneman, et al. (1991), Arkes (1991), Schwarz (2000), Kahneman (2003 & 2012), Tetlock (2005), Mintz (2004a, 2004b, 2007), DellaVigna (2009), Ardanaz, et al. (2013) Hafner-Burton, et al. (2013 & 2017), Montibeller & von Winterfeldt (2015), Thaler & Ganser (2015)

It is not yet entirely clear how these different heuristics and biases interact, but it may nonetheless be useful to hypothesise. This study aims to use cognitive bias to explain IR phenomena considered bizarre through the lens of IR theory. In the following chapters, I will outline these bizarre events in more detail. Further, I will systematically examine the main literature in BIR to show that, through its focus on cognitive bias as a concept (and core variable in thinking of IR), it has an important role to play in expanding our understanding of the international political world.

The above concepts, which outline the main cognitive biases and heuristics observable in IR, help greatly to create a framework for this task. In noting the existence and nature of these biases, we can begin to make the case that they are in fact influential, and that they transform strange phenomena into relatively comprehensible events.

7. Conclusion

This chapter begins by making the case that human beings are more than merely rational. To characterise the human cognitive process thusly is limiting and incorrect. At the very least, human beings do not conform to expected utility models of choice and other rationalistic choice models. Human beings lack the time, energy, capacity or information to conform to these models. Instead, it is more pertinent to conceptualise human beings as reliably irrational, where we make systematic errors in judgement and information processing.

I outline some of the key theoretical literature, which argues that people are influenced more strongly by emotive, psychological and biological factors than they are by strategic gain. Individual preferences and beliefs matter greatly to the human brain, often more than questions of net gains. Under this behavioural approach, a 2-stage model of cognition has been widely accepted. This model distinguishes between intuitive, instantaneous processes and rational, deliberative ones. The patterns of biases and heuristics that characterise human thinking vary predictably depending on which mode of cognition one uses in a given scenario.

Other theories break down errors in thinking into categories depending on the biological foundations of the errors. Other theories categorise these errors by viewing them as the culmination of cognitive processes. This theoretical foundation then justified the conceptual framework I created to describe the human cognitive process. This framework described a

process that included: i) reception of information, which included the manner in which information is accumulated and stored ii) perception of information, which referred to the way information was analysed and processed and iii) decision, which denoted the act of weighing and measuring the processed evidence to conclude about which the correct course of action is. The Decision step in my conceptualisation was also shown to be heavily influenced by the decision-makers preferences and beliefs.

The proposed framework for cognitive bias represents an important phenomenological claim. We do not, intuitively, have complete access to the external world, upon experiencing it we immediately store information in an imperfect way and change it. However, the manner in which we interpret reality are not – in practice – infinite. Rather, they are predictable across people. We can use this insight to i) better predict and understand human behaviour and ii) reconcile varying views based on their (almost inevitable) commonalities. I believe that these two insights are central to the aims of BIR, and that the fact that my conceptual framework of cognitive bias reveals how and why, at a conceptual level.

Chapter 3

Distilling BIR:

An Integrative Review on the Processes and Applications of the Field

1.1 Introduction

This chapter will embark on an integrative review that will analyse the key research focus, methods, findings and frameworks of six texts that have laid the foundations of Behavioural International Relations and brought it into the sphere of relevancy. The six texts it analyses are: "*Judgement under Uncertainty: Heuristics and Biases"* by Tversky & Kahneman (1974), "*Perception and Misperception in International Politics" by Robert Jervis (1976)*, "*Behavioural IR as a Subfield of International Relations"* by Mintz (2007), "*The Behavioural Revolution and International Relations"* by Hafner-Burton et al. (2013), "*The Micro-Foundations of International Relations Theory: Psychology and Behavioural Economics"* by Stein (2017) and "*Political Psychology in International Relations: Beyond the Paradigms"* by Kertzer & Tingley (2018).

The aim of this chapter is twofold. Firstly, it aims to distil the most influential literature in the field into a comprehensive and comprehensible space, one that gives at once a clear vision and deep understanding of the field of Behavioural International Relations. This will be achieved through the review that follows. The review will break down each piece of literature into component parts:

i) Study Summary (brief outline of study type, methodology and experiment design, if applicable).
ii) Key findings and implications of these findings for IR.
iii) Areas for future research.

The second aim of this chapter is to form a discussion which integrates the main findings of each identified piece of literature into a conceptual toolkit which will help scholars to view IR events and political leaders with enhanced perspective. The second section of the chapter will embark on a critical discussion on the value of the findings of the literature for explaining

IR phenomena like those listed in the previous chapter and how to incorporate these findings into models and theories that explain and predict events in IR. This approach adds the variable of cognitive bias to considerations of how decision-makers act and react, and so should create insightful explanations for certain outcomes in the international political realm. This discussion will promote the main case I am making with this research: The methods and practices of BIR, and the concept of cognitive bias in particular, has the potential to explain events in IR which have been inadequately explained by the dominant paradigms associated with the field.

1.2 A Short Note on Sampling Techniques

This type of review has some significant shortcomings. Since it is qualitative in nature, there is no sampling technique that I could use that would be randomised. Indeed, randomisation in selection would be unfavourable, since the data would not all be concerned with the issue here at hand – namely, cognitive bias in International Relations. One might argue then, that for our aims here, a purposive sample would be in order. Where a purposive sample would aim to collect "crucial research, [where] the data is meant to contribute to a better understanding of a theoretical framework" (Ilker Etikan et al., 2016:2), a convenience sample would be chosen for characteristics such as "easy accessibility, geographical proximity, or the willingness to participate" (Ilker Etikan et al., 2016:2).

I think that a critique of this study that claims that the sample of texts I have used is a convenience one falls flat. The methodology section clearly outlines the selection criteria that I used, and only a study outside the scope of this one (perhaps a more exhaustive doctoral degree for example) may find more relevant texts than the ones I have identified. Even here, I remain sceptical, but admit that a more experienced researcher (one with years of experience) may be aware of less cited work that is of high relevance but whose lack of keywords or citations may ensure it fails to be present on the online search engines which made up the majority of my literature search.

Quite apart from this point, however, if the sample selection does lapse into convenience sampling, the critique falls flat anyway. I aim to present an integrative review which calls for scholars to consider the value which the methodological and theoretical foundations of Behaviouralism and BIR add to IR theory. I aim to do this by considering that cognitive bias is

an important explanatory variable in completing the puzzle of international relations. A convenience sample, if it substantiates this point, is no less effective than a thoroughly purposive one.

2. Judgment Under Uncertainty: Heuristics and Biases (Tversky & Kahneman, 1974)

i. Study Summary

This ground-laying work has very little need for introduction. It revolutionised behavioural thinking in many different fields, and the authors have brought the notions of heuristics and biases into the vocabulary of academics and the wider population alike. There are perhaps no more qualified authors on the topic. I include it in my literature review precisely for these reasons. The authors have been actively involved in a variety of fields, most notably psychology and economics, and yet the broad implications of their research mean this seminal text qualifies – under the criteria I identified earlier – as highly relevant for BIR. In fact, work by Khaneman and Tversky informs a wide range of topics in behavioural studies across disciplines. A cornerstone of BIR is its ability and willingness to learn from and incorporate relevant research from a wide range of fields into IR. I hope that this becomes progressively clearer throughout this study. Kahneman and Tversky are two authors who illustrate indelibly that cross-discipline research can be fruitful and engaging.

The 1974 study by Tversky & Kahneman involved experimental research. By this I mean simply that independent variables were manipulated, participants were randomly assigned to different stages of this manipulation and this response to the different manipulations were observed. The researchers describe 3 heuristics in making judgements under uncertainty. I will briefly outline the experiment design and then move on the implications of these experiments in the next section.

The authors tested the representativeness heuristic in the following manner:

Participants were given the descriptions of different individuals. These personality descriptions were said to be sampled at random from a group of 100 people comprising of engineers and lawyers.

An example of one description: "Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail."

Participants were then asked, for each description akin to the one above, to judge the probability that the person (Steve) was an engineer instead of a lawyer. They were also told that the sample group from which the description had been drawn consisted of varying numbers of lawyers and engineers. In one case, participants were told that the group consisted of 70 engineers and 30 lawyers. In the second case, participants were told that the group consisted of 30 engineers and 70 lawyers. Given the statistical data, we would expect that the first group would consistently rate the odds of Steve being an engineer as more likely. In the second, we would expect that the Steve is considered more likely to be a lawyer. However, it was shown that – regardless of the stated conditions – participants did not come to this conclusion. Instead, what appears to have happened is that people evaluated the probability that a particular description was an engineer or lawyer based on the representativeness of that description to the stereotype of an engineer or a lawyer. Prior probabilities appear not to have influenced the participant's assessments.

The authors tested the availability heuristic in the following manner:

A few different instances showed the predictable errors displayed by participants. First, participants were asked to listen to a list of names compromised of 50% men and 50% women. They were then asked to assess whether the list contained the names of more men or more women. Different lists contained different names, where in some lists, the men were relatively more famous than the women. In the other list the opposite was true. Participants stated the number of male or female names were larger when the names of that particular sex were, relatively, more famous. In another example, a group of participants were asked if a randomly selected word from an English text either i) begins with the letter 'r' or ii) has the letter 'r' as its third letter. Most people judge the likelihood of the former to be higher and the latter lower.

The authors tested for the adjustment and anchoring heuristic in the following manner:

Two groups of high school students were asked to estimate a mathematical product within 5 seconds. One group was asked to assess the product: 8 x 7 x 6 x 5 x 4 x 3 x 2 x 1 while the

other was asked to assess 1 x 2 x 3 x 4 x 5 x 6 x 7 x 8. The group that was asked to assess the first estimated a higher median value than the group who were asked to assess the second. While adjustments were far lower than the true value in both cases, they were close in the first case because the result of the first few steps of multiplication are larger in the first example (Tversky and Kahneman, 1974:1128).

In a second example, a wheel of fortune was spun. Participants were then asked to judge the percentage of African countries that were members of the United Nations. When the arbitrary value on the wheel was 10, median estimates were 25 and when the arbitrary value showed 65, median estimates were 45 (Tversky and Kahneman, 1974:1128).

ii. Key Findings and Implications for IR.

This article examined biases that are the result of judgemental heuristics. The experiment focused on cognitive biases which were not affected by motivational factors, since subjects were rewarded for correct answers. Thus, the errors in judgement analysed in this article stem from the cognitive processes we assume are logical, rational and analytic (Tversky and Kahneman, 1974:1130).

The errors in judgement the authors identify include those individuals with training in statistics, that is to say, professionals. These unconscious biases apply even to the intuitive mind of experienced scholars. Elementary errors such as the gambler's fallacy may be avoided by trained professionals, but more complex concepts evade the intuition of even those experienced in statistics. We should, therefore, be extremely mindful and wary of these biases, especially when they apply to IR.

Decision-makers appear to display three important heuristics when making judgements under uncertainty. Firstly, participants in the experiment illustrated the representativeness heuristic. Participants evaluated the probability that object A belongs to class B. The participants evaluated the probability that A belonged to B based on the degree to which A represented B. The implication here is far-reaching. The authors demonstrate that participants are insensitive to predictability and make decision under an illusion of validity. Simply put, decision-makers look for highly consistent patterns when predicting outcomes. Descriptions of a company that case it in a good light – but provide no information about the profit or future profit – are likely to cause participants to rate future profit as higher. There

is consistency between a positive description and positive profit, but this consistency should not figure in true estimates of future profitability, since the description may be biased, incorrect or irrelevant.

The human mind does not intuitively note this. People are also more confident in their predictions when they note highly consistent patters, for example, participants were more confident rating the future marks of a student who achieved all B's on the report card compared to a student with many A's and C's. In an ironic statistical phenomenon, consistent patterns are often observed when the inputs are highly redundant or correlated, which significantly lowers the predictive power of the model. Thus, people tend to be more confident in their predictions even as the statistical likelihood of their accuracy falls (Tversky and Kahneman, 1974:1126). In the realm of IR, we should be very careful when political advisors are very confident. It is important to assess the roots of this confidence, and to be critical of what we see as consistent patterns in what is likely to be a random event.

Secondly, they display the availability heuristic. That is, they judge likelihood of an event or the frequency of a particular class based on the relative ease with which they can bring instances of that event to mind. The implications for IR are obvious, and numerous. We are likely to overrate the probabilities of events we can recall easily, and so are elite politicians. For example, terrorist attacks are perceived as relatively more likely than they, in reality, are. It takes a concerted effort to form policies that focus on issues more likely to cause death and difficulty for the population (Tversky and Kahneman, 1974:1131).

Governments are apt to inefficiently distribute resources towards defence budgets compared to awareness campaigns about the risks of heart disease, the importance of frequent screening for cancer or programmes for the suicidal, all of which are in the top ten leading causes of death in the United States (National Center for Health Statistics, 2017).

Thirdly, the authors identified an anchoring and adjustment heuristic. The basic finding here is that participants use an initial value (which may or may not be arbitrary) to anchor their estimation. The anchor is then adjusted to get a final answer, this adjustment is usually insufficient. Again, the implications are numerous. A strategy for overcoming the bias would include discarding the use of anchors, using more than one, adjusting estimates by larger amounts, making greater use of statistical analyses for assessing likelihood of outcomes and

spending time and effort gaining as much relevant information as possible before taking a course of action (Tversky and Kahneman, 1974:1131).

The authors also note that heuristics are not themselves biases. Heuristics can, and often are, very effective. They shorten the mental process and make it effortless, in short, they are economical. However, they do lead to errors, and these errors seem predictable.

iii. Areas for Future Research

The areas for future research from this seminal work have largely been explored. The heuristics and biases the authors identify have been tested further and proved in other settings. Of course, the authors were working in the field of psychology and economics, and the applications of thee biases to instances unique to IR have not been extensively explored. As will become clear throughout this integrative review, other scholars have attempted to experiment with political elites to determine if these biases still exist and that level and to what extent.

While BIR is in the process of incorporating these concepts into its own research agenda, it is moving slowly. One of the biggest questions is this: What can political leaders do to mitigate these cognitive biases? The answer to this question is one that is important to answer, and the benefits to society, domestically and internationally, are potentially vast.

3. Perception and Misperception in International Politics (Jervis, 1976)

i. Study Summary

Jervis' work is a cornerstone for Behavioural IR. His work was the first comprehensive and successful attempt to integrate psychological analysis into mainstream IR theory. Jervis attempts to answer questions like: "What are the causes and consequences of misperceptions?" (Jervis, 1976:3) and "How are beliefs about politics and images of other actors formed and altered?" (Jervis, 1976:3). Jervis, in his introduction, presents one of the key aims of his book. That is, to demonstrate that misperceptions on the part of decision-makers should not be treated as mere accidents, or extraordinary events. Jervis' opening claim is that "we can find both misperceptions that are common to diverse kinds of people and important differences in perceptions that can be explained without delving too deeply into individuals' psyches" (Jervis, 1976:3).

Part one of the analysis is entitled "The Setting", and lays the foundation for Jervis' work. In chapter one, the author outlines some of imperfections in the decision-making process, and provides a requisite account of the psychological approaches and attempts to assimilate them into the language of International Relations. Chapter one deals with perception and the level of analysis problem. Insofar as we consider the analysis of the decision-making process to be important in IR, it is necessary to explore the variables that are situated between how actors see the world and how they statesmen will act in the world. Jervis asserts, throughout the book, that it is the case that heterogeneous preferences for policies are traceable to differences in perception of the environment. Jervis (1976:15) presents four levels of analysis. They are 1) decision-making, 2) bureaucracy, 3) the nature of the state and domestic policies and 4) the international environment. If we only needed the last three, non-decision-making levels, if all we required to predict a decision was to know the objectives of the involved actors, their bureaucratic systems and their state's interests, then there would be no need for analysing the decision-making process itself. However, we find that we have the information from the last three levels of analysis, we cannot predict outcomes in IR.

In chapter two, Jervis examines "how actors attempt to understand and predict the actions of others" (Zinnes, 1978:793). Jervis successfully presents a framework relating to the intentions of international actors, through which one can use historical data (past behaviour) to infer future actions. The basic premise here, backed up by many historical examples, is that the analysis of past behaviour, in itself, is insufficient to predict future conduct. The inference process that Jervis lays out requires statesmen to analyse a number of factors. These include the characteristics of the state, the "constellation of forces, beliefs and goals…together with the external stimuli the state is likely to face" (Jervis, 1976: 33). However, when analysing past actions, we must "try to understand why the other acted as he did [since] no bit of behaviour is self-explanatory or has only one plausible implication for the actor's future conduct" (Jervis, 1976:33).

In chapter three, the author presents critique of the two chief models which are used to explain the importance of both intentions and perceptions in the decision process. Jervis attempts to show that the theorists of the psychological approach and the deterrence theorists do not satisfactorily address how to determine and predict "the intentions of the other states in the system" (Jervis, 1976:9).

Jervis argues that deterrence theory is based on the fear of aggressor states. According to this argument, the aggressor state will constantly test the strength of the status quo power, acting aggressively and opportunistically if it senses weakness in either resolve or capability. Thus, it is necessary for status quo powers to check the aggressor state with displays of power, or face greater difficulty in convincing the aggressor state of its resolve in the future. Where "the deterrers worry that aggressors will underestimate the resolve of the defenders, the spiral theorists believe that each side will overestimate the hostility of the other" (Jervis, 1976:84).

For these "spiral theorists", without a global sovereign, each state needs its own military to protect it. However, since there is never a guarantee of everlasting peace, the amassing of military might in the present constitutes a threat in the future. This leads to a spiral, where one state seeks to build up a military more powerful than those of its adversary. Perception plays a part in the model in the sense that own states view their actions as defensive, but the same actions by other states as aggressive. Jervis (1976:75) states that: "The inability to

recognize that one's own actions could be seen as menacing and the concomitant belief that the other's hostility can only be explained by its aggressiveness help explain how conflicts can easily expand beyond that which an analysis of the objective situation would indicate is necessary."

Thus, spiral and deterrence theory stand in direct contradiction. Jervis (1976:84) states it most succinctly and it worth quoting at length once again. "Policies that flow from deterrence theory (e.g. development of potent and flexible armed forces; a willingness to fight for issues of low intrinsic value; avoidance of any appearance of weakness) are just those that, according to the spiral model, are most apt to heighten tensions and create illusory incompatibility. And the behaviour advocated by the spiral theorists (attempts to reassure the other side of one's nonaggressiveness, the avoidance of provocations, the undertaking of unilateral initiatives) would, according to deterrence theory, be likely to lead an aggressor to doubt the state's willingness to resist."

Ultimately, Jervis recommends that politicians implement policies that present asymmetric payoffs. That is to say, where a state's assumption about its adversary is correct, there should be high payoffs, but if a state is incorrect in its assessment of the other, there should be tolerable consequences. This would mitigate the fact that we can never know with certainty what others are thinking and consequently, whether to operate under a deterrence model or a spiral model.

Part two of Jervis' work begins with chapter four and ends at chapter seven, it is aptly named "Processes of Perception". In chapter four, Jervis discusses rational cognitive consistency and irrational cognitive consistency, and examines different sources of cognitive biases. The author suggests that rational decision-making requires one to ignore or alter incoming information. Jervis notes that while some assumptions do not deal with all the information, they remain rational and relatively unbiased. For example, "we tend to believe that countries we like do things we like, support goals we favour, and oppose countries that we oppose. We tend to think that countries that our enemies make proposals that would harm us, work against the interests of our friends, and aid our opponents" (Jervis,1976: 118). Such assumptions do not analyse all the data, but remain useful and rationally consistent.

Irrational consistency involves a biased consistency, one which (usually negatively) affects one's worldview and decision-making. Generalising that while a policy succeeds in one area, it must therefore succeed in another is an example of irrational consistency (Zinnes, 1978).

Chapter five considers the context in which information is received. It discusses the process of image formation. Since it is difficult to dislodge images of others and perceptions of different instances, it is vital to come to terms with how these images and perceptions come to be.

Continuing with this line of thought, chapter six gives an account of how the influences of international history, the agent's domestic political system and her non-political training all affect her perceptions and reactions in variety of situations and with other actors. Jervis core argument is that actors seek to construct frameworks through which they can understand complex new situations. One such framework is historical analogy, especially when it has related policy decisions or crises. Unfortunately, the actual links between the present event and the historical one become somewhat irrelevant, and consequently decision-makers do not see the entire picture when analysing a policy problem. The eventual outcome may be that a decision-maker implements a sub-optimal policy because it was previously successful or that the decision-maker ignores the optimal policy because the policy was historically unsuccessful.

Policy-makers make other errors in judgement too. Jervis (1976: 281-282) notes that "there is often little reason why those events that provide analogies should in fact be the best guides to the future…because outcomes are learned without careful attention to details of causation, lessons are superficial and overgeneralized…decision-makers do not examine a variety of analogies before selecting the one that they believe sheds the most light on their situation." Jervis (1976:235) refers to this as "inappropriate learning".

Chapter seven examines the extent to which and the mechanisms through which attitude change occurs. It indicates the possibility that information contrary to pre-existing perceptions can alter those perceptions.

Chapter eight is the first chapter in the third section of Jervis' work. The theme of the third section is "common misperceptions", and chapter eight deals with "perceptions of centralization" (Jervis, 1976:319). Chapter eight posits that we have a tendency to assign

order to disorder. That is to say, when analysing events in hindsight, we tend to construct a coherent narrative to events that are largely the result of chance. In so doing we "see the behaviour of others as more centralized, planned, and coordinated than it is" (Jervis, 1976: 319). The implications for International Relations are significant, and can lead to the misperception of competence, guile and threat in an adversary. Jervis (1976:338) notes that "taking the other side's behaviour as the product of a centralized actor with integrated values, inferring the plan that generated this behaviour, and projecting this pattern into the future will be misleading if the behaviour was the result of shifting internal bargaining, ad hoc decisions, and uncoordinated actions."

Chapter nine discusses the degree to which an actor overestimates its "importance as influence or target" (Jervis, 1976:343). Jervis proposes that, in the absence of full information about the intentions of another actor (an adversary, say) one will consider oneself the target of that adversary's policies, and thus overestimate the degree to which the adversary is a threat. Jervis (1976: 349) states that "when others' actions hurt or threaten the perceiver, he is apt to overestimate the degree to which the behaviour was a produce of internal forces and was aimed at harming him."

Chapter ten contemplates the influence of desires and fears on perceptions. It also analyses the effects of wishful thinking in International Relations. The wishful thinking bias is present where "subjects overguess the frequency of desired outcomes and underguess the frequency of undesired ones" (Jervis, 1976: 362). Jervis argues that, by and large, international political decision-makers are not wishful thinkers. Realism's influence on the political mind probably trains such decision-makers to err on the side of caution and not wishful thinking. In the discussion that follows the analysis of this and the other core texts of this study, I will expand on the notion of wishful thinking in international politics.

Chapter eleven deals with cognitive dissonance in International Relations. Cognitive dissonance is pervasive in this area, where decisions have significant risks and rewards. Indeed, the global scale of the decisions which leaders must make increase their will to view their actions as useful and productive. Jervis (1976:406) notes the central contribution of cognitive dissonance theory: "people seek to justify their own behaviour – to reassure themselves that they have made the best possible use of all the information they had or

should have had, to believe that they have not used their resources foolishly, to see that their actions are commendable and consistent."

Chapter twelve is the last chapter and the only chapter under section four – which is entitled "In Lieu of Conclusions" (Jervis, 1976:409). The last sections of Jervis' work deals with how we might plausibly minimise misperceptions among leaders in IR. Jervis' policy recommendations draw from all of his earlier analyses to justify themselves. Jervis notes some fairly apparent solutions (is not the wisest wisdom apparent?). He suggests that decision-makers ensure their beliefs and values are expressed explicitly (especially on important issues), that leaders should play devil's advocate (form contrary images) and use third parties to assess the effectiveness of policy. Jervis (1976:423) notes that the "most obvious safeguard is for decision-makers to take account of the ways in which the processes of perception lead to common errors."

Jervis' study is still one of the "most extensive analysis of the perceptual dimensions of foreign policy" (Eldridge, 1977:1108). His work was one of the first and most influential in the attempt to bridge the rift between foreign policy work and the study of perceptions.

ii. Key Findings and Implications for IR

Jervis' core thesis is not novel, but the combination of the scope and depth of his analysis does provide a wholly compelling case for his claims. His core thesis is clear: "foreign policy decision-makers' perceptions of their environment and of other actors often diverge from reality with important consequences for their and others' subsequent actions" (Eldridge, 1977:1107).

Jervis notes the key role of different perceptive and practical imperfections in our ability and will to collect and data and process information. The lack of complete date, the inability to conclusively determine the existence and strength of causation and the paradoxical perceptual biases of the scholarship itself all play a role in forming perceptions in IR (Pentony, 2005). Jervis believes that IR ignores psychology at significant cost. Some common categories of misperceptions are outlined by Jervis, including the tendency of people towards wishful thinking (Jervis, 1976:356). Additionally, Jervis discusses issues including cognitive dissonance, and the tendency for people to restructure their personal narratives so as to

remove a degree of individual responsibility and to avoid evidence that runs contrary to their beliefs (Jervis, 1976:406).

Jervis argues that existing systems of analysis do not address puzzles at an individual decision-making level. A focus on the external or international system, the domestic scheme or the bureaucratic level of analysis are all argued to come up short vis-à-vis a focus on the effects of the perceptions of individuals. Jervis also notes the impossibility of processing all available information and the role of heuristics in decision as a crucial part of rational reasoning. Jervis focuses specifically on the human need for consistency, and many of his findings can be traced back to this observation. Jervis incorporates a variety of disciplinary research into his work in order to conclude that perception is a crucial component of the decision-making process, and that our need to form coherent narratives about the story of our actions drives much of our behaviour (Jervis, 1976: 392-404).

Despite writing the original text in 1976, Jervis' insights are fundamental. Work in political psychology has certainly increased in the last half a century, but this work has not replaced the work of Jervis or made it redundant. If anything, it has made it more relevant. In the preface to the second edition of *Perception and Misperception*, Robert Jervis admits that if he were to rewrite the book, he might focus more on the role of emotions, motivated biases and the power of expectations and needs in our decision-making analyses (Jervis, 2017: lxxxviii).

On a fundamental level, Jervis managed to bring the individual level of analysis to the forefront of academic minds in International Relations. While other work has been done in this area, none of it resonates quite as strongly as it does in *Perception and Misperception*. Rose McDermott, commenting at a roundtable event dedicated to discussing Jervis' work comments: "it proved seemingly insurmountable to apply concepts developed around individual drives, motives and feelings to understand the more complicated, interactive and structurally constrained actions of nation-states. All of this changed with the publication of Perception and Misperception, which opened up an entirely new way of studying the influence of the individual on state action" (Fujii, 2017: 7).

iii. Areas for Future Research

Jervis creates a model for the sort of interdisciplinary research that Behavioural International Relations aims to produce. The integration of multiple disciplines and the wide collection of documentation that the author manages to accumulate is most impressive. Jervis uses historical examples from an array of different time horizons and incorporates these examples with illustrations and explanations from fields as diverse as physics and philosophy. While Jervis work provides an exceptional example of the use and possibility of such an approach, his work is far from finished. These connections are still largely underdeveloped and future research would do well to emulate his approach, and build on the model and insights he provides.

4. Behavioural IR as a Subfield of International Relations (Alex Mintz, 2007)

i. Study Summary

Mintz outlines the central ideas and assumptions of the subfield of Behavioural International Relations. He sets out the key questions that BIR addresses, its main methods and the rationale for the existence of the field. The essay also addresses the possibility that Behavioural IR can facilitate the scientific study of decision-making and the historical justification outlining the need for a field focusing on the issues of BIR (Mintz, 2007:157).

The method the author uses is probably best described as a literature review. The author assorts a range of literature in order to identify anomalies from "the traditional analytic, rational, expected utility model of choice." The accumulation of these behavioural studies forms the foundations from which we can justify the importance of BIR. Mintz (2007:158) identifies those biases which have been studied through a process of reviewing them and extracting their essential findings. His essay also identifies the main contributors to ideas and examples which fit into the aims of BIR. Mintz (2007:159-160) uses his review of the literature to characterise BIR as a field in its own right, to position it as a unique method as well as imbue it with a conceptual entity distinct from other fields.

This piece of literature might well be the first one that identifies BIR has a distinct field. It represents the first academic acknowledgement of the need and utility of the field, and the first evidence for the claim. This text consists a particular kind of literature review and critical discussion. It is a series of essays that makes the case for the necessity of BIR. In it, Mintz points to previous research indicating the existence of "numerous anomalies from the traditional analytic, rational, expected utility model of choice" (Mintz, 2007:157). These include but are not limited to: framing effects, what Mintz himself termed the polyheuristic bias, emotional bias, loss aversion, wishful thinking and groupthink.

Mintz lays the framework for asserting that leaders are biased decision-makers, as are we all. Most decisions, in complex environments are subject to cognitive bias. Mintz argues that the explanatory power of IR is improved by incorporating these insights into its analysis. He notes that IR already deals with the problem of non-rational thinking. Constructivists, for

example, occupy a space in IR theory that focuses on societal and personal values as influencers. BIR simply expands this. Mintz points out numerous studies that show how cognitive bias can influence IR by affecting the judgements of political elites. He cites the example of the invasion of Iraq to illustrate his point (Mintz, 2007:159).

Once he has established that the concepts of cognitive bias and non-rational decision-making do indeed play a significant role in IR, Mintz constructs a picture of the field for us. He notes that behavioural IR, in line with other behavioural sciences, is a "process-oriented, quasi-rational, involves 'satisficing,' and the study of framing, all in a bounded rational environment." Moving beyond these behavioural traditions, however, BIR also deals with emotion, perception, belief, culture and many other factors. Mintz goes on to point out that the behavioural literature is varied, it is far from homogenous in its use of systems, theories and frameworks. Study in the behavioural literature, whether in the political sciences or other sciences, ranges widely in focus. Some studies focus on the way leaders acquire and are presented with information, and the way this may influence decision-making. Other focuses include the study of the impact of individual personality, operational codes, roles, beliefs, preferences and cultural norms on the decision-making process. Many of these areas of focus coincide with the research agenda of behavioural game theory, which informs much of the literature in BIR.

James (2007:162), writing in the publication Mintz edited, makes some practical suggestions for the field of BIR. James suggests some directions that scholars might move towards in the interest of promoting and developing BIR as an independent subfield. He notes the important contributions BIR promises to produce, including its work in understanding preference formation and advancing scientific progress in International Relations and the social sciences more broadly. James concludes by affirming the viability of BIR as a subfield in IR and acknowledges its potential to being a significant contributor to IR in the future.

Walker (2007:166), as part of the forum discussing the viability of BIR (edited by Mintz) as a subfield of IR makes an additional contribution to the discussion. Walker argues that "a combination of favourable methodological developments and historical conditions has combined to create a bright future for Behavioural IR as a subfield." Walker further notes the characteristic of BIR to focus on the micro-foundational relationship/s between individuals,

between states and between individuals and state or non-state groups. This focus seems to exist within the realm of bounded rationality and on the beliefs, preferences and psychological variables of these actors. Walker also addresses the fact that Behavioural International Relations has existed for over half a century. However, there are new developments, in technology, psychology and other factors, that make it cheaper and easier to map and analyse political elites with greater experimental validity.

ii. Key Findings and Implications for IR

Mintz's essay on *Behavioural IR as a Subfield of IR* catalyses a process that is vital for the future of BIR. His essay extracts the most fundamental pillars of this emerging field and articulates them extremely succinctly.

The author identifies six characteristics of BIR and identifies the relevant actors, concepts, methods and questions of interest in BIR (Mintz, 2007:159-160). Behavioural studies are not new to the political sciences, however, there is no collection of concepts that outlines BIR in such a collected way. It is arguable that Mintz's essay represents the first meaningful conceptual synthesis of the field of BIR. His framework informs much of the scope of my study. By identifying the characteristics and other relevant variables, Mintz narrows the scope of study in BIR and makes its goals clear. He notes the assumptions of BIR, an important step towards establishing it as a fledgling field. Some of these assumptions include: Political elites often make suboptimal strategic decisions, leaders focus on narrow sets of information and use cognitive shortcuts to process information, BIR seeks to understand process as well as outcome and the assumption that bias influences decision-making at an individual, group or state level. These fundamental tenets of BIR act as a guide that defines what exactly should be considered part of the field.

Actors in BIR range from state, business and military leaders to nations themselves, domestic agencies and the public. The levels of analysis of BIR vary with the relevant actors a study examines. A study may focus on the individual, group or organisational level. Variables like the nature and severity of a threat, the nature of the decision-making unit, or the relative difference in leadership style might drive the preferred level of analysis. Concepts in BIR include that which influence decision-making. They range from beliefs, emotions and preferences to information acquisition patterns, culture and groupthink. Specifically, issues

such as framing effects, ambiguity and inequality aversion and anchoring are of interest to the field (Mintz, 2007:158). Methods in BIR include, but are not necessarily limited to: practical experiments, computational modelling, game theoretic experiments on behaviour, surveys and elite interviews and statistical analyses. Different methods vary in their uses and shortcomings. In line with the philosophy of the behavioural sciences, experiments and data driven knowledge is vital to the field. The validity of many BIR findings rests on the empirical evidence and experimental robustness of the studies within the field.

Questions of interest in Behavioural International Relations represent a curious amalgamation of behavioural questions and political ones. They include such issues such as war – its initiation, escalation and termination, peace, deterrence, international trade and finance, environmental concerns, negotiation and the decision-making of terrorists and political elites (Mintz, 2007:160).

The questions of interest of BIR are significant in reducing the scope of the field. While some findings in the behavioural sciences apply to issues of human decision-making in general, the environments of these decisions are important to BIR. The field seeks to answer questions relevant to international society. It is clear that Mintz calls for the utility of a behavioural approach to be rediscovered in International Relations, to understand and study issues of international importance in novel, non-paradigmatic ways.

James (2007), wrote an essay in response to Mintz (2007) and gave some practical suggestions for moving BIR forward. One approach he suggests is a that of “Systemism”. This framework is necessary, James (2007:164) posits, because the numerous components that make up the field are problematic. For example, how do emotions, cognitive bias and heuristics interact? When do they conflict or compound the effects of the other?

James makes the argument for an organising principle that links these different processes and outcomes causally. It is necessary for integrated causal frameworks to be established if we are to logically expand research in BIR. Systemism, as James describes it, is a possible solution to this issue. The essay is not long enough to move much beyond the theoretical possibility of this perspective. Essentially, the approach of Systemism is a theoretical approach. James argues that it “goes beyond macro-macro and micro-micro theorizing to include hybrid linkages as well: macro-micro and micro-macro” (Mintz, 2007:164). The core

idea of such a theory would be to create links across systems where these links create a causal and hierarchical structure between concepts. In addition to having an organising principle for the concepts of BIR, an inclusive ontology and the compiling of evidence will be extremely valuable in creating a body of useful and productive research field.

iii. Areas for Future Research

Behavioural IR as a Subfield of International Relations is perhaps the first codified work regarding BIR as a distinct and useful field. Research has been conducted in the field since 2007, but not much. There exists a massive array of potential areas for future research. The scope of the options is daunting, and without a conceptual and theoretical framework for developing the field, we may be in danger of failing to create a congruent system within which to understand the issues of BIR.

Future research could be conducted in any of the areas that Mintz (2007) outlines. These include the exposure of leaders and other political elites to framing effects and negative political information. It also includes phenomena such as emotional bias, loss aversion, wishful thinking and the "shooting from the hip" (Mintz, 2007:158) bias. More specifically, the rigour of these findings could be established in a political setting, since many of the findings have unrepresentative samples or apply to economic contexts. This will become more possible, as James (2007:164) notes, with the growth of communication technology, which will allow researchers to collect data from political elites.

In order to consolidate the accumulation of knowledge in the field, it would be extremely productive to settle on a categorisation of the concepts of BIR. It is my hope that my contribution to this discussion provides an example of how to approach this important goal.

5. The Behavioural Revolution and International Relations (Hafner-Burton et al., 2017)

i. Study Summary

The authors discuss the behavioural revolution by standing it in opposition to the assumptions of rational choice. They note areas in which these two approaches complement each other and approaches which seem to represent more fundamental disagreements about the nature and behaviour of actors. Behavioural work introduces insights into beliefs and preferences which could feasibly be integrated into rational choice models, with updated assumptions. Other research indicates variability in decision-making due to emotion, among other things, which more radically undermines rational choice theory.

The method the authors use may be classified as a narrative review. They do not specify their classification for their own work, but it may be justifiably inferred. Hafner-Burton et al. (2017) identify numerous anomalies in IR, ranging from the Iraq invasion, to international political economy and foreign policy analysis. The authors note the growing significance of these insights, as well as the previous contributions to the field, beginning in the 1950s. The study by the authors aims to incorporate behavioural insights into the field of IR more fully. The authors hope that, by focusing on actor heterogeneity, it promises to spawn more empirically realistic (and accurate) models of decision-making.

According to the authors, behavioural work in IR appears to emphasise the effects of individual heterogeneity on the decision-making process. This is a promising development and is likely to improve our understanding of the political landscape, its often-strange realities and the role of institutional design on the decision-maker. There is considerable debate as to how behavioural studies in IR will move forward. There remains an open question: Will behavioural work influence the field more effectively than it did in the last century, or will it open up greater rifts between deductive theory and empirically driven research?

ii. Key Findings and Implications for IR

The authors acknowledge the key work in behavioural economics that paved the way for the behavioural revolution in IR. They note that Kahneman and Tversky (1974 & 1979) made

some pivotal findings. These include such insights including: i) the tendency for people to use heuristics which under or overestimate probabilities and lead to less-than-maximum expected utility outcomes, ii) Individuals asses utility in terms of net gain from a reference point, and values assigned to losses or gains are asymmetric (Hafner-Burton et al., 2017: 14). Insights like these led to the development of a theory of cognition particular to behavioural studies, but this spawned another issue.

At the birth of the behavioural revolution, scholars were preoccupied with identifying universal deviations from rational choice. As Hafner-Burton et al. (2017:9) note, a "drawback of this approach was a laundry list of biases, but with the corresponding difficulty of knowing when such deviations would arise, with what magnitude, and whether they were in fact constant across agents." The authors note that the project of systematically analysing these biases has driven progress in the field. The authors note how non-standard assumptions, i.e. assumptions that differ from rational choice models, are of great interest to contemporary scholars. The authors break down non-standard assumptions into three categories: i) Nonstandard preferences, ii) Nonstandard beliefs and ii) Nonstandard decision-making procedures (Hafner-Burton et al., 2017: 14). These categories break down the "laundry list" of biases into manageable and understandable groups. The authors show how these biases, listed under the three categories, have been studied in IR and the possible extensions researchers could explore. These core drivers of actor heterogeneity act as a useful frame from which to approach these cognitive biases.

Hafner-Burton et al. (2017) collect and synthesise a multitude of literature in the behavioural fields and identify 4 major sources of actor heterogeneity. Though, as they state, this is not an exhaustive list, it opens up a research agenda that may integrate studies in the field more closely and allow behavioural studies to diffuse into International Relations significantly. The authors posit that differences in resolve (essentially political will), differences in social preferences, level of public absorption and use of information and experience constitute 4 key sources of actor heterogeneity.

The authors illustrate that new behavioural work is widening our dimensions of heterogeneity and assorting their causes in systematic ways. In addition, these differences in decision-making processes are more empirically grounded and their psychological

foundations are being noted with increasing appreciation (Hafner-Burton et al., 2017: 13-15). Some of these insights are compatible, even synergistic, with existing IR literature. Other insights require more fundamental rethinking of rational choice theory, its validity and application (Hafner-Burton et al., 2017: 22).

Two cautions are given by the authors. The emerging behavioural literature must contend with two important critiques. Firstly, surveys or experiments are often used to identify the existence and effects of psychology on beliefs, preferences and strategic thinking. Some of these surveys or experiments focus on decision-making political elites, but most use convenience samples such as university students (Hafner-Burton et al., 2017: 21-22).

This is not a new research problem, but the focus of IR is on political elites, who may differ from other strategic thinkers. Additionally, it is often impossible to replicate the stakes of international political decisions in a lab environment. These problems are, of course, not unique to International Relations, but the robustness of findings needs to be established, and the external validity of findings reiterated, if IR is to become – as I hope – a well from which other fields might draw for important insights and effective methodological approaches.

iii. Areas for Future Research

The authors note a key issue explored, but not solved, by the behavioural movement in IR. The issue is the aggregation issue. How exactly do we move from individual to collective decision-making? How are biases mitigated or exacerbated by institutions? Do nation-states have preferences that are determinable and compatible with predictive models, as one approach suggests, or should we focus on psychological analyses that deal with decision-making processes? These questions are important to ask because they will inform the research agenda of BIR for years to come.

Congruent with these paths for future research, are the issues discussed above. The robustness and external validity of previous and future findings needs to be as well-established as possible. The task of reiterating and streamlining the survey and experimental work of Behavioural IR falls to the scholars of IR (Hafner-Burton et al., 2017: 23).

Other areas for future research indicated by the authors are numerous. They include issues such as the differences in time-weighted behaviour between democrats – who favour the short term, and autocrat – who favour the long term. Other areas for future research range

from bias towards familiar, but suboptimal policies, climate change negotiations, foreign aid and many more. The behavioural revolution in International Relations is relatively new, and there exists a plethora of research opportunities, especially when one specialises research into the realm of the policymakers, leaders and other elite samples.

As I've noted previously, some insights from behavioural studies need not undermine already established models of rational choice for scholars. Indeed, research that incorporates new insights about the human decision-maker might simply adjust certain assumptions given the existing parameters. In so doing, behaviour-focused work can be relatively easily incorporated into IR with relevancy and effectiveness.

6. The Micro-Foundations of International Relations Theory: Psychology and Behavioural Economics (Stein 2017)

i. Study Summary

Stein (2017) traces a new wave of scholarship in International Relations that draws deeply from psychology and behavioural economics. The author notes that previous contributions by behaviouralism did not diffuse broadly into the field of International Relations (Stein, 2017:249), this despite contributing significantly to theory and research. The author hopes to move the debate beyond the rational choice vs behaviouralism paradigm. Instead, having spent the last few decades identifying the scope conditions of rationality, researchers have an opportunity for determining a more empirically justified baseline for human behaviour. As the author so concisely notes, "What is anomalous or puzzling can only be determined against a baseline, and the more empirically grounded the theories that specify the baselines, the better the choice of puzzles and the more productive the research agenda" (Stein, 2017:259).

The author uses a narrative literature review method, bringing together studies in behavioural fields to provide a useful roadmap of the field. The authors beings by noting some of the obstacles to diffusion for behavioural work in the latter half of the 20th century. These obstacles included the lack of proper scope conditions for how and when agents would deviate from expected behaviour. As a result, deductive theory, which was not trouble by these experimental difficulties, was preferred by a large portion of scholars working in IR.

ii. Key Findings and Implications for IR

The author notes that there is a difference between the behavioural movement that is taking place today and the movement of the latter 20th century (Stein, 2017:257). Where historical behaviouralism challenged rational choice and attempted to discredit its assumptions, contemporary behaviouralism seeks to establish scope conditions for cognitive biases and the thresholds before which these biases might be reversed. The question is not: "Are people rational?" but rather "under what conditions are decision makers likely to be rational and when are they likely to behave in ways that behavioral theories expect?" (Stein, 2007: 250).

Stein also summarises the contributions of prospect theory to questions of international security. The author notes the role of loss aversion, probability weighting, framing effects, reference dependence and many other factors in the decision-making process (Stein, 2017:251). The author makes note of promising methodological approaches which incorporate these behavioural insights into practical applications. These approaches include a robust test of loss aversion in deterrence outcomes. This approach includes a careful process of identifying the reference points of decision makers before the event in question. Thus, it is possible to establish the net gains decision-makers face relative to the identified reference point and assess whether an outcome was in fact predictable, even after the fact. Another approach the author notes is the approach of using the "hedonic tone" (Stein, 2017:253) of an issue to posit a natural frame. These frames affect a decision-makers weighting of losses and gains through loss aversion.

Of course, the behavioural movement faces some significant challenge. The author notes two core challenges. Firstly, the issue of external validity is important. Experimental results inside of a laboratory is not necessarily applicable outside of the experimental setting. Another issue is that of aggregation and disaggregation (Stein, 2017:255). Assuming that an experiment in behavioural studies achieves external validity, there remains the issue of applying findings to levels of analysis that differ from the experimental setting. How, if at all, do individuals influence group-level processes? Are cognitive biases, and to what extent, exacerbated or lessened by organisations? The literature indicates a movement away from these polarised questions. Instead, as with the rational versus irrational debate, are moving to ask questions about the scope conditions for groupthink, and the nature of the relationship between groups and individuals under certain conditions. This promises to be a more fruitful approach (Stein, 2017:255).

Stein (2017:257) notes that scholars are focusing on concepts of social contagion, loyalty and consensus. Other authors explain the relationship between individual-level bias and the circumstances under which these biases are mitigated or worsened. Other solutions include assigning individual-level characteristics to states, or the focusing on the aggregation process as a filter for the number, direction and consequences of choices.

The author concludes by noting the contemporary benefits of the internet for experimental surveys and experiments. Studies of specific samples like elites will be vital to validating generalisations from experiments in the laboratory. Large-N studies will likewise remain vitally important to the field, in order to test the extent of psychological concepts on leaders' behaviour (Stein, 2017:259).

The micro-foundations of behavioural IR are being established. Authors like Rathbun, Kertzer and Paradis (2017) are identifying the nature of strategic rationality, noting the psychological traits that drive it. These include a high regard for self, a motivational desire for knowledge. Reciprocity and fairness have also been found to play an important role in bargaining. Authors like Bayram (2017) also identify the role of social values and self-identification within a social group in complying with international law.

iii. Areas for Future Research

The intersection between psychology and rational choice presents a range of opportunities for future research. The work that established a human pattern of thinking that showed the limitations of rational choice was useful, but it is now a largely stale debate. The next generation of scholarship can use behavioural game theory, psychology and economics to create systems and models that uncover the causes and nature of escalatory behaviour and collective action problems.

These types of analyses will begin to establish the micro-foundations of strategic behaviour, which will specify the scope conditions for rational and non-rational behaviour. Scholars have, and should continue to, deal with the question of who is strategically rational and under which circumstances.

Another area for future research concerns establishing thresholds for the effect of psychological-behavioural effects on strategic rationality (Stein, 2017:259). By drawing a line that identifies when psychological factors affect the rational outcome, we can begin to mitigate and predict outcomes more effectively. This type of research extends to numerous cognitive biases, and the thresholds of these cognitive biases will no doubt be different. Additionally, the behavioural movement introduces a doorway into the bargaining literature (Stein, 2017:258). Cooperation and preferences viewed through a broader lens of social

preferences and collective identities promises to be helpful in explaining compliance in the absence of enforcement and bargaining between states.

7. Political Psychology in International Relations: Beyond the Paradigms (Kertzer & Tingley, 2018)

i. Study Summary

The authors make the case for the emergence of behavioural research and psychology in international relations. They do so by examining the state of the field in a review essay that analyses four years of study classifications at one major International Relations journal. In so doing, they give a "data-driven snapshot" (Kertzer & Tingley, 2018:1) of the type of psychological questions IR scholars are and are not investigating. The authors make use of radar plots in order to map the author-selected classifications, methodologies and other substantive content of IR scholarship in general against political psychology within IR. The authors analysed all of the submitted manuscripts at the International Studies Quarterly (ISQ) journal, noting that it is "the flagship journal of the International Studies Association" (Kertzer & Tingley, 2018:5) and is representative of the myriad of IR scholarship. The time period in question begins in 2013 through to the beginning of 2017.

The authors conduct this study in order to show the significance of questions that move beyond traditional IR paradigms. They focus on growing areas of research, including the increasing importance of biological, evolutionary and psychological lenses in the political sciences. The authors argue that these developments in IR mirror broader changes in the discipline, which are arguably for the best. The increasingly diverse methods and questions that are being applied to IR is opening up the literature to insightful analyses – moving beyond errors in decision-making to a theory of cognition deeply important to IR. The authors further caution scholars to import psychological insights critically, and a call to craft a recognisable ontological and epistemological framework in order to ensure the coherence of IR scholarship.

ii. Key Findings and Implications for IR

The authors create a useful, and empirically driven, map of the scholarly work in IR. The data indicates that IR scholars in general are focusing heavily on statistics and case studies as their preferred method. These two methods dominate the research significantly, indicating a lack of experimental and survey analysis, as well as a lack of genealogical, ethnographical, archival and interpretive methods (Kertzer & Tingley, 2018:7-8).

It seems that political psychologists within IR utilise a broader set of methodological tools, especially with regards to experimental research and surveys. Almost a third of political psychologists used experiments compared to only four percent of ISQ submissions on the whole (Kertzer & Tingley, 2018:6). These findings are not particularly surprising, but again, they do call for a critical examination of how we teach IR. If IR is to become a more empirically driven field, there is a case for moving away from the dominant paradigms we tell students are so important to the field. We need to develop and articulate the identity we inhabit as IR scholars rather than the one we imagine.

We've covered the methodology that IR scholars are using, what about the topics they are covering? Political psychologists have a natural interest in behavioural issues. Behaviour is nothing but psychology, at the fundamental level, and so it is interesting for the field of BIR to become acquainted with the questions of political psychology, and to communicate with them. Political psychology is fairly diverse in its research agenda, but some patterns are clear. Questions of interstate conflict, IR theory, diplomacy, race, gender and social movements are all important to political psychologists. Most important of all, are questions of political parties, elections and public opinion, which dominate the research over the last half a decade.

Questions of international institutions, international law, foreign direct investment, monetary policy and global governance are not well represented by political psychology. The authors put it like this: "psychological research in IR remains focused on conflict rather than cooperation, and on behaviour rather than institutions" (Kertzer & Tingley, 2018:6). I think this is the dichotomy that BIR can help to destroy. It is necessary and useful to categorise, and we should not do away with categorisation in general. However, I believe it is erroneous, and not particularly useful, to distinguish between behaviour-focused work and institution-

focused work. Are institutions devoid of behavioural bias? Do individuals make decisions within an institution? Is an institution nothing but a group of biased individuals attempting to make decide on a course of action? There is certainly an argument to be made that by partitioning scholars into fields in the way that is evident – i.e. political psychologists handle individuals and other IR scholars will deal with institutions – we stand to ignore the clear links between research questions and entire fields of study. BIR hopes to be a part of the solution.

The authors go on to note the six directions political psychologists are exploring that are different from those they have explored in the past. These include: i) a shift to non-rational assumptions about behaviour, influenced by behavioural insights, ii) A greater interest in mass political behaviour and political opinion, iii) experiments involving elites, iv) interest in international political economy (IPE) and, v) most distinctively, growing interest in genetic, evolutionary and biological approaches to studying IR (Kertzer & Tingley, 2018:3-4).

iii. Areas for Future Research

IR is lagging in exports. The field is not highly relevant in other fields, but it could be. The academic economy is a rich one, and IR scholars – especially those conducting psychological work – are in a better position to contribute to other fields now than ever before. With increasing focus on experimental work, the insights IR scholars gain provide fields such as economics, psychology and even biology, are increasing.

Over and above those areas for future research which have already been identified in the previous sections, there is perhaps a crucial area that we – as IR scholars – must focus on. I mentioned the false dichotomy that is often drawn between the individual and the institution, or the individual-level analysis versus the group-level analysis. Of course, groups are not merely the sum of their parts, but they are the function of them. To view them as distinct, as I briefly argued, is to miscalculate, and the results hinder the accumulation of knowledge. What seems lacking most in the arena of behavioural research in international relations at the moment is the communication between two groups of scholars. The first group are normative, theory-centred scholars, concerned with global governance, institutional cooperation and international organisation. The second group are the political psychologists and behaviouralists, and they are focusing on questions of conflict and public elections.

I think a significant contribution to IR scholarship may yet come from within BIR. What is necessary for this contribution is quite simple in theory, though it will be far from simple to accomplish in practice. There is a need for a bridge to be built that connects the individual with the institutional (Kertzer & Tingley, 2018:12). The nature and complexities of experimental research make it convenient to analyse individuals, who are often not organised into groups. If scholars can conduct experiments about decision-making process within institutions, formulate conceptual, methodological, ontological and epistemological frameworks that analyse the links, feedback loops and communicative failures of institutions, we may be able to integrate two levels-of-analysis and begin to remedy some institutional failures. The best frame for organisations, I would argue, is a collection of individuals. How and when this collection affects already-established biases is the question we need to address.

8.1 Integrating Findings

Earlier in the study, I identified the distinguishing characteristics of an integrative review. One of these characteristics was the use of a narrative synthesis to create an insightful perspective on the relevant issues. In this section, therefore, I aim to analyse how the aforementioned seminal texts relate to one another, and tease out the evolution of thoughts, concepts, methods and relevance of each text.

I will attempt to present the evolution of thinking around cognition, perception and bias in International Relations and, by necessity, other related fields. My hope is that by making use of the insights we are now familiar with from the previous discussion, I can condense and synthesise the crucial findings, developments, assumptions and frameworks which form the academic backbone for Behavioural International Relations and notions of cognitive bias in the international political world.

8.2 Narrative, Analysis and Synthesis

The story of Behaviouralism in International Relations is incomprehensible without first attending to its relation with other academic fields. One such example is the effect that work in general theories of psychology and economics has had on the Political Sciences and, by extension, International Relations. Behaviouralists, working in IR, have used the fields of economics, psychology, history and philosophy to inform their approaches and concepts when dealing with political phenomena. The 1950s and 1960s saw the expectant wave of the behavioural approach rise and subsequently come crashing down. Dahl (1961) gives us some perspective as to why this wave crashed as early as it did. Dahl (1961:763) notes that "The behavioural approach, in fact, is rather like the Loch Ness monster: one can say with considerable confidence what is not, but it is difficult to say what it is." The use of broad, numerous and diverse terms – from all manner of fields – confused the debate and the identity of these behavioural scholars. Shortly after Dahl decried the lack of a cohesive 'behavioural identity', Rosenau (1965) admitted deep frustration by the work of behavioural political scientists. He notes his hope that the behaviouralists would "demonstrate the virtues of a broad-gauged, interdisciplinary approach to international phenomena" (Rosenau 1965:509). His hopes were frustrated by the lack of cross-disciplinary adaptation that would need to accompany insights from other fields in order to make them relevant to the political world. Indeed, while behavioural scientists worked in their respective disciplines, they superficially noted the "multidimensionality of conflict" (Rosenau, 1965:521) and requested consideration for their work in the political sphere, never infusing their work with political relevance or communicating to the political scientist about how he might adapt these insights for his purposes. Rosenau ends his discussion on the links between behaviouralism and international phenomena by calling not for the production of new and untested ideas, or the promotion of values, but rather for the "never-ending pursuit of comprehension" between scholars of different disciplines (Rosenau, 1965:521).

His wishes were realised when three scholars reignited the fervour for the behavioural sciences with two field-defining texts. The first two were Kahneman and Tversky. Although Kahneman would earn his Nobel prize after the passing of his co-author, Amos Tversky, and would earn it for their work entitled *Prospect Theory: An Analysis of Decision under Risk*

(1979), Kahneman and Tversky influenced thinking in the political realm at least six years prior to this publication. In the above section, I speak about the seminal text of Tversky and Kahneman (at least, the one that is seminal to International Relations) entitled *Judgment under Uncertainty: Heuristics and Biases (1974).* Thomas Christensen, presenting an introduction at the International Security Studies Forum (ISSF), where the works of Jervis were discussed, makes an important point. He notes that "Jervis is enamoured of general theories of…human psychology, like that of Tversky and Kahneman" (Fujii,2017:2). It is difficult to imagine Jervis' work being as influential without the priming work of Tversky and Kahneman. Of course, Jervis' work stands alone as a defining text for IR and particularly BIR, yet it is important to note the influence that the two psychologists undoubtedly had on his publications.

If one considers it for any length of time, one will notice the enormous debt that all behavioural scientists owe to the field of psychology. Freud, after all, popularised the notion of the unconscious mind (Fujii, 2017). Although he did not develop it to the point that we now understand it, the notion of a mind at once our own and yet often outside of our conscious control is central to the work of Kahneman and Tversky and all behaviouralists. This is clear when we seek to identify the predictors of behaviour, which often diverge from an individual's stated aims, preferences and values. This reality is only possible in a world where we find it impossible to translate what we want into how we act. The reasons for this discord arise from the individual, but are not clear even to him. The mind, in this view is not the sovereign of the person, but rather a powerful actor, unable to dominate yet able to influence. It is, rather fittingly, like a state in this sense.

How has the behavioural approach in International Relations developed? Is it still like the Loch Ness monster, or have we managed to define it, use it and integrate it usefully into our study of the international world? As always, the answer is not simple, a frustrating yes and no. Kahneman and Tversky reignited the embers of a slow burning fire, and certainly helped to catalyse the behavioural approach across disciplines. *Judgment under Uncertainty: Heuristics and Biases* (1974) describes three heuristics that are used when making judgements under conditions of uncertainty. They illustrate the tendency of decision-makers

to use the heuristics of representativeness, availability and adjustment. The authors show a proof of concept. The human machine is not as simple as we might hope, it errs in weighing decisions, and it is biased. But, like the work before it, this particular text, as well as *Prospect Theory: An Analysis of Decision under Risk* (1979), does not integrate (it does not try to) perfectly into the realm of the political. Fortunately, Jervis notes the gap and fills it extraordinarily well.

In fact, in some respects, Jervis' work was so definitive – or so it seemed at the time of its writing – that those "who read it felt like there was little that could be added" (Fujii, 2017: 6). Jervis managed to do most of what Rosenau, in 1965, declared would be necessary for the continued relevance of behavioural studies in the political world. Little more than a modicum of work from psychology had managed to find its realisation in the subfield of IR. Jervis managed to move the direction of behavioural studies in IR beyond the study of leaders, into the realm of the nation-state – where complicated structural constraints and interactivity all underpin behaviour. Jervis' work on *Perception and Misperception* was able to move beyond Freudian psychoanalysis and "examine ways systematic and predictable biases in the human decision-making apparatus could influence leaders" (Fujii, 2017: 6). While Jervis' work plays a foundational role in BIR, he does not intend to provide a unified theory of cognitive bias or perception in IR. Instead, as Jervis notes in the preface to the new edition of *Perception and Misperception (xxi),* he presents an inductive study, which draws from cognitive and social psychology and then attempts to match these theories with his reading of diplomatic history.

In the absence of a unified theory, research in these areas continued slowly in the years after Jervis' work. Research cemented and expanded the case that Jervis and those before him had presented, that decision-makers acted in ways that differed importantly from the expected utility model of choice. Importantly, however, this research cemented previous observations in the political realm by illustrating their applications in the field. Levy (1997, 2003), Mintz (2004a), Nincic (1997), Forman and Selly (2001) and Janis (1982) all showed the existence of behavioural anomalies in the political realm, among foreign policymakers and state leaders. The inclusion of *Behavioural IR as a Subfield of International Relations,* edited

by Alex Mintz (2007) reiterates the case for moving towards a unified theory of cognition in IR. In fact, the case is strong enough to propose an entire subfield, BIR, which hopes to focus on cognitive variables in its analysis, incorporate those methods and theories which suit its needs (but perhaps not the needs of all scholars in the Political Sciences) and build an updated theory of individual and state behaviour. Mintz (2007) seals his place in this analysis by providing some measurable and foundational characteristics for defining BIR as a subfield. If we are to develop anything approximating a unified theory, it will become increasingly important to define the bounds of scholarship. Additionally, though I have advocated frequently throughout this study that interdisciplinary scholarship is central to the type of work that behaviouralists undertake, it is important to define the boundaries of study. The alternative is a field mired by inconsistencies and scholarship that contributes little to the common research agenda. Indeed, it threatens the very existence of a common research agenda to begin with. Progress is stifled by too much creativity as well as too little. The conclusion of Mintz's work (2007:165) is that BIR must needs adopt a "systemist frame of reference and investigate all four of the possible kinds of linkages, from macro-macro downward through micro-micro." Further, there is a place for both case studies and statistical analyses in BIR, which will need to come together in an inclusive environment of theorising and evidence compilation to confirm individual hypotheses or test wider "frameworks with interconnected propositions" (Mintz, 2007:165).

Mintz (2007) lays foundational work, but how have his ideas developed, and have they come to fruition? Hafner-Burton et al. (2017) make strides in this direction. The authors present the standard position of Behaviouralism as an alternative to the rational choice approach. Noting the drawbacks of approaches that list a number of biases without classification or frameworks to conceptualise, explain and mitigate them. Instead of this approach, Hafner-Burton et al. (2017) create a framework for dealing with the notion of cognitive biases by identifying three categories of non-standard assumptions about the behaviour of actors. They also identify and categorise four sources of actor heterogeneity. What makes this text worth including in my analysis, and what distinguishes it from other texts in the field? In *The Behavioral Revolution and International Relations (2017),* Hafner-Burton et al. manage to fit different sections of the behavioural puzzle together. Where previous studies identified, for

example, one source of actor heterogeneity and another may have dealt with non-standard beliefs, these authors synthesised and categorised such findings in an accessible and methodological format. The result is an evolution of the unified theory that we are seeking in Behavioural IR. The work is something other scholars can communicate with in a meaningful and structured way, which is an important step in the pursuit of understanding in this area.

Hafner-Burton et al. (2017) also develop and illustrate the common terminology that emerges from a reading of relevant texts in IR, Political Psychology and Behavioural Economics. The concepts of "nonstandard preferences, beliefs and decision-making" (Hafner-Burton et al., 2017:6) allow us to grapple with the theory that follows from the observations of behavioural work. If actors do not conform to the expected utility model of choice, if their decisions are not rational in the cost-benefit sense, how then should we define their utility functions? What, in fact, are their preferences, and how are these shared preferences spread among decision-makers? These are the types of questions that a model of nonstandard preferences, beliefs and procedures can answer. Extensive work has been done to show how actors diverge from expected decisions. Yet there is comparatively little work in the area of defining how actors *do* make decisions, as opposed to how they *do not* make them. Table 1 on page 14 of Hafner-Burton et al.'s work (2017) is an example of how we can begin to categorise different anomalous behaviour and hypothesise about how actors make decisions under different conditions. For example, actors weigh losses more heavily than gains, they prefer the present to the future and they prefer certainty to ambiguity. These are preferences that are identified under the nonstandard preferences categorisation. This type of identification is useful, necessary and insightful. However, it does lead naturally to another problem. Decision-makers certainly display nonstandard behaviour, but only under some conditions. Undoubtedly, in areas of uncertainty, or high stress, actors make errors in assessment. However, in purely probabilistic games, actors can also be shown to act according to the expected utility model of choice. It seems that there is no one set of preferences, beliefs and decision-making processes that captures the entirety of the human experience. This is not really a surprise, since an entirely complete theory would be a theory of everything. However, it does present the following issue: Under what conditions do human

beings act according to standard models of choice and under what conditions do they act under nonstandard models? This is the problem that Stein (2017) deals with in *The Micro-Foundations of International Relations Theory: Psychology and Behavioural Economics.* It is precisely because Stein deals with this issue that the work is included in the review, since it represents an answer to what I consider a significant issue in Behavioural IR. Without specifying the conditions under which actors act irrationally (according to nonstandard models of choice), there is no possibility of theoretical consistency, explanatory or predictive power.

Another key challenge for the development of the field is the levels of analysis issue. It is unclear to what extent individual-level effects aggregate out to group-level effects. It is also unclear under which conditions (if any) and to what extent (if any) individual biases are mitigated or exacerbated by group settings. This is a necessary and interesting expansion of work in the field. These issues, among others, are of major concern to Stein (2017), and as such the text represents a movement in consensus within the field. The text summaries some of the key concepts in BIR, implicit in which is the understanding that there exists a common idea of what these concepts mean to the behavioural scientists and that they are, for the most part, facts rather than theories. The concepts of loss aversion, framing effects and time inconsistency are examples of strong building blocks with which the behavioural scientist can construct a sturdier model of decision-making in the international sphere.

Over the last half a century, behavioural studies in IR have sought to question assumptions of human behaviour. Their insights have, in large part, become part of the common understanding of human nature. Kertzer and Tingley (2018) give a more empirically grounded insight into the state of IR scholarship. While this text does not delve deep into theoretical developments and the evolution of concepts in political psychology and Behaviouralism, it gives us an objective data point from which to tie together a coherent narrative. Kertzer and Tingley (2018) lend credence to the story I have told here above. I have indicated that there has been an adoption of nonstandard assumptions about human behaviour, that there has been a movement towards an individual level of analysis and that there is increasing focus on experimentation with relevant subjects (i.e. political elites). The evidence I have given has

involved the narrative synthesis of behaviourally-focused literature in relevant fields. However, the case for the narrative I have given is strengthened by the study of Kertzer and Tingley (2018), since it provides empirical evidence that the story presented is accurate. Behaviouralism is becoming more prominent in the Political Sciences, as are its insights and contributions to the academic literature. This suggests that the subfield of BIR, still underdeveloped, enjoys fertile soil from which to sprout, and that it could aid those scholars looking to participate in the work that it encourages, but who are uncertain about the particular field that they occupy.

9. Conclusion

This chapter reviewed six texts situated in or directly related to the field of Behavioural International Relations. It distilled the most influential behavioural work in IR by giving an overview of the relevant studies. This included the methods, findings, implications and areas for future research of each of the relevant texts. The aim of this summary of information was to show the relevancy and state of the field of BIR, the questions with which it engages, it theoretical and methodological philosophy and its ability to give us useful insights into the international arena.

The chapter then engaged in a narrative synthesis that attempted to illustrate how these core texts communicate. It attempted to tease out the evolution of thoughts, concepts, methods and assumptions of each seminal text in order to facilitate the discussion. Behaviouralism in Political Science has, in some cases been contrasted with rational choice. In other cases, it has played a complementary role to rational choice models, by updating models of actor preference, beliefs and the cognitive process. In the area of emotions, more fundamental differences are at play between these two schools of thoughts. Since the first behavioural revolution, more than half a century ago, significant contributions to theory and research have been made. In moving forwards with this research, scholars have begun to identify the scope conditions for both standard and nonstandard assumptions about actor behaviour. The experimental nature of contemporary scholarship gives some scholars hope that this new behavioural revolution will disperse more widely into IR. Other scholars in IR are wary of the rifts that could open up between “deductively valid theories and empirical

research in international relations" (Hafner-Burton et al., 2017). It is my belief, grounded in the evidence I have here presented, that the behavioural scholarship in International Relations over the last fifty years has been insightful and useful. Future research will need to synthesise a more comprehensive and unified theory of behaviour in IR. If it does, it may constitute some of the most relevant work political scientists have conducted in recent years.

Chapter 4

Misbehaviour or Misunderstanding?

1. Introduction

This chapter posits that what has been perplexing to IR scholars and often labelled bizarre or irrational is merely misunderstanding. By referring to events in this way, casting them in the light of 'misbehaviour', we lay the foundations for dismissing them all too easily. Dominant paradigms in IR fail to note the role of individual cognitive bias in decision-making and their related consequences in the international world. Dominant IR theories, like Realism, Liberalism or Institutionalism fail to explain certain phenomena for a few reasons. Their focus on states as the key unit of analysis causes them to ignore valuable insights that help to understand the real world. The world of ideas that scholars often find themselves sparring in is somewhat removed from the realities of international politics. What is often cast, by the dominant IR paradigms, as bizarre behaviour might simply be the result of the human decision-making process in action.

This chapter notes some of the major shortcomings of traditional IR theory. This chapter argues that International Relations is framed – when it is taught in the classroom – in terms of the dominant paradigms or "-isms". However, scholars of IR evidence that they research in areas that are not characterised by these paradigms. This disconnect probably contributes to the fact that IR literature is not incorporated into other fields. The chapter argues that the failure of IR to diffuse into other fields is, in part, a result of the failure to incorporate the concept of cognitive bias (among other concepts) into the research agenda and IR theoretical toolkit. In attempt to present a balanced cased, the chapter also addresses some of the possible responses the traditional approaches to IR might have for behavioural scholars.

The chapter then demonstrates that there are certain phenomena that are poorly explained by traditional IR theory. Specifically, the issues of (non)compliance in IR, beliefs in policymaking, the invasion of Iraq, the forecasting inaccuracy of political experts and climate change negotiations are examined. This chapter also hopes to demonstrate that the gaps in

BIR theory has prompted the model of cognition that I developed earlier in this study. The existence of useful insights and applications, paired with the lack of a unified theory of cognition in BIR presents an opportunity for a more cohesive framework for approaching these issues. This chapter does so by making good use of Table 3 on page 39

2. Dominant IR Paradigms: A Short Exposition

System-level variables remain the focus of International Relations theory – though some schools of thought, e.g. Constructivism, focus on individual-level variables. Despite the move away (in most of the recent literature at least) from easily defined paradigmatic analyses, the core tenets of traditional IR theory remain in place (Slaughter, 2011). What exactly are some of the tenets? (Walt, 1998)

i) The duality of anarchy and sovereignty remain central to the field. The international world is anarchical (there is no clear leader state with authority over other states).

ii) State or non-state (private groups or organisations) are the main actors in IR. Thus, these (usually) group entities become the units of analysis. Individuals may be important in Liberalism but are usually seen as a political force only when they are collected in groups and pursue the group interest.

iii) The distribution of power is central to state interaction

iv) Variables that determine how states or non-state actors often include: military power, material interests, ideological beliefs, the pursuit of peace or some other reasonable and definable goal.

v) Actors are basically rational, i.e.) they attempt to maximise their own gains – they may have preferences but these preferences/norms merely form part of the mediation of self-interested pursuits, and they are socially (not biologically/psychologically) constructed. Even liberalism argues that co-operation is in everyone's interest and thus conforms to the rational agent model.

In the discussion below, I will do my best to briefly and accurately describe the foundational elements of International Relations theory. Cognisant of the clichéd and stale lectures on the dominant IR paradigms, and simultaneously aware of the need to give an account of them in order to point out some areas of weakness, I will focus only on those points that are pivotal.

2.1 Realism

Realism posits the nation-state as the principle actor in IR, where individuals and organisations exist, but with far more limited influence. States are assumed to be unitary actors, unfragmented, speaking and acting with a single voice. Insofar as rationality encompasses the pursuit of national self-interest, nations are assumed to be rational actors. The state of anarchy encompasses the environment in which all the above takes place, where no single state is in charge of the others (Waltz, 1979). Nation states attempt to navigate this competitive environment in order to protect their existence and prosperity.

Realism conceptualises agents as self-interested, egoistic and power-seeking. Realists also believe that this behaviour is predictable and repetitive over time. Power and deception are crucial tools for the conduct of foreign policy in the realist account (Elman, 2007). Realism has certainly evolved from its inception in Thucydides' history of the Peloponnesian War and Machiavelli's *The Prince.* Kenneth Waltz's 'neorealism' emphasised structural explanations for behaviour rather than philosophical presuppositions about human nature. The international anarchic structure, the fact of power differences between states and the will to further state interest became the core of investigations by Realists in the latter half of the 20th century.

It is arguable that IR theory is most often used in the world of policymaking, compared with the other dominant IR paradigms. This, at least, is the realist claim, and echoes Machiavelli's desire to inform world leaders on statecraft and the necessary interplay between Lion and Fox. Though Realism claims to reflect the reality of the international political world, and despite its contribution of valuable insights into the human condition and the condition of international politics, it has received wide criticism. Realism failed to predict the end of the Cold War, for example, because Realism overlooks the potential for ordinary citizens to rebel and gradually overthrow existing power structures. The end of the Cold War illustrated the possibility of a more optimistic vision of international relations. Critics also note the narrow focus realists have on the state, ignoring the importance of individuals and organisations on issues that go beyond mere state interest. Realism has also been criticised for its amoral position, where leaders legitimise the use of violence to further their self-interest (Antunes & Camisao, 2017).

2.2 Liberalism

Liberalism is arguably a less cohesive body of theory than Realism. Liberalism's core insight is that national characteristics and differences matter for their international relations. As opposed to Realism's characterisation of states, where the goals and international behaviour of states is essentially the same across actors, liberal scholars emphasise the role of state heterogeneity in IR (Slaughter, 2011).

The moral principle that ensures the right of individuals to life, liberty and property defines Liberalism. Political systems, in this view, are formed in order to protect the rights of its citizens. However, national institutions are not the only necessary conditions for liberty. Militaristic foreign policy troubles liberal theorists, and – while necessary for preventing military conquest of other states – military power should be limited through institutional intervention or other measures. Liberal democracy is a mechanism by which to allow a state to build up military might without subverting individual liberty, as is the division of political power among different arms of government. The arguments of democratic peace theory are strong. There is empirical evidence that liberal democracies do not go to war with one another, since they are characterised by internal restraints on power and see one another as legitimate and non-threatening. There is, however, evidence that democracies are likely to be aggressive to non-democracies – illustrated by the 2003 Iraqi war.

States may not necessarily be primary actors in the liberal view. Individuals and private groups are often seen as fundamental actors in world politics, they are referred to as non-sate actors. States, in the liberal view, are representatives of the dominant subset of domestic society, and serve their interests. The way these interests are configured determines state behaviour, so ideological beliefs and commercial interests figure as important issues alongside the survival of the state. Institutions and norms both exist to limit the power of states. Internationally, organisations such as the United Nations foster cooperation and impose costs on those who violate international agreements. Economics and economic interdependence are seen to be a powerful (dis)incentive for states to conform to international norms and agreements. Liberal theory emphasises different variables in its analysis of international behaviour, and though it moves away from the state as the only important actor, the theory still focuses on institutions, groups and collections of

interests as core drivers. There is not a significant emphasis on elite politicians and their decision-making process in liberalism. Individuals are important mainly when their views, votes and preferences are aggregated across society, rather than as influential in their own right. Contemporary liberal theorists argue that absolute gains outweigh relative gains for states, and that cooperation and trust is likely to increase welfare for all. This implicitly assumes that actors in IR weigh decisions based on some analysis of the utility of certain decisions and act to maximise (or at least increase) this utility. In this sense, liberal theory can be said to be rationalistic (Meiser, 2017).

2.3 The English School

English school theory is built around a trio of concepts: the international system, international society and world society. It claims to offer a middle ground between the oft-opposed schools of Realism and Liberalism. Hedley Bull (1997:9-10) describes the formation of the international system: "when two or more states have sufficient contact between them, and have sufficient impact on one another's decisions to cause them to behave as parts of a whole." The international system is conditioned by the state of anarchy in which these states exist. The internal society that emerges in this view does so when a common set of rules, norms and institutions binds like-minded states together. Additionally, world society supersedes international society, where the ultimate unit is not the state but the individual. In this way, the English school attempts to create an overarching theory which encapsulates both the view of individual as primary actor and state as primary actor in IR (Stivachtis, 2017:29)

The distinction between an international system and an international society is often questioned within the English school. It is also unclear as to where the line between the two concepts should be drawn. There is relative acceptance that "an international system constitutes a weak/thin form of an international society" (Stivachtis, 2017: 34). The English school also highlights the important change in international society, from a world of conflict pre-1945 to a world of relative global peace thereafter. Accompanying this international change is the emphasis on human rights and the perception of the world as a single, global economy and a global economy. By embedding common ideas through technological media,

it is increasingly possible for political and economic elites and ordinary citizens to share and protect common ideas, which is a stabilising force in international society.

The English school focuses on the study of history in order to understand contemporary world politics. For the English school, the balance of power is not the only important variable, in order to understand a state, it is important to understand a state's history, as well as the particular threats and motives of each state (Stivachtis, 2017:28). Domestic politics, norms and ideologies are all important to English school theorists. They do not seek to create general explanatory models, however, instead, nuance and context is vital to interpretation. The English school is also critical of rationalist assumptions in International Relations, emphasising the important and complex role of society on shaping behaviour in the international political order (Slaughter, 2011).

2.4 Constructivism

The birth of Constructivism was, in a significant way, connected with the failure of both Liberalism and Realism to predict the end of the Cold War. Where many other dominant paradigms in IR fail to account for the role of the individual in the international political world, constructivist theorists note that influential individuals play a vital role in shaping and reshaping the nature of IR through their interactions with other individuals and actors (Theys, 2017:36-38).

Constructivism is so named because its basic theoretical presupposition is that what we know, and can know, of the world is socially constructed. Meaning, or ideational structure, is vital to interpreting and perceiving the world. There is no true objectivity, in this view, because we are not capable – as human beings – of internalising information without affecting it. Social context in the constructivist view is pivotal in the view we hold of reality. Reality is, in many ways, being formed in the present. Ideas and beliefs distort, or create, the reality we experience.

The behaviour of states is closely related with perceptions of friends and enemies (us versus them), fairness, justice and virtue. Though some Constructivists would agree with the supposition that states are self-interested, they may argue that this self-interest is complicated, and ultimately subjective. Survival, power and wealth, for example, are likely not the only goals that the state pursues. Norms, not only consequences, inform the

decisions that actors in IR take. For example, even if annexing a neighbouring state may allow a powerful one to gain an important strategic resource, this move violates important international norms that precedes an otherwise clear cost-benefit analysis. Constructivists note the role of Non-Governmental Organisations (NGOs) or corporations as influencers and creators of norms (Slaughter, 2011).

Where the English school, Realism and Liberalism focus on the distribution of material power, resources and status, Constructivism emphasises ideas, identities and norms in determining state behaviour. There is room in Constructivism for fruitful dialogue with Behavioural International Relations, and it is possible that BIR can build upon its foundations to further delve into the role of the human mind in altering and interpreting events in the international world.

2.5 Critical Approaches

There exists a vast array of theoretical approaches in IR, and it is not the objective of this study to detail all of them. As such, and despite being aware that grouping all so-called 'critical' approaches together does involve a fair amount of conflation, I nonetheless think it is a necessary and useful tactic for their exposition here.

Critical approaches in IR include such perspectives as Marxism and Feminism, among others. Critical approaches are joined in their rejection of the underlying epistemology of other IR theories as well as the constructions of power and the state that Realism, Liberalism and even Constructivism take for granted (Slaughter, 2011).

Marxism, for example, look past state to state relations and focus on a fundamental system of global class relations. The dialectic between capitalists and the working class underpin state relations, and overlooking the interests of these two groups and their interaction would constitute a severe and dangerous misrepresentation in the Marxist view. Feminism focuses on gender as a key variable in IR, which has led to notions of human (individual) security – which moves beyond state (military) security. Issues in IR, such as war, affect family life and social relations as well as borders and issues of resource or power distribution.

Critical approaches shift the focus in IR from power and wealth towards emancipation, human freedom and other such concerns. People, in this perspective, should be placed at

the centre of political discourse. Moral issues, the identification of political alternatives and the historical processes that shape both the past and – potentially – the future are all issues worth examining from these perspectives (Ferreira, 2017:49-53).

3. Challenges to IR Theory

The so-called "Great Debates" are often used to narrate the shifting nature of ideas in IR. Regardless of the technical accuracy of framing the narrative this way, the idea nevertheless permeates and influences the field. All fields are subject to the ebb and flow of ideas, where some drift out to sea while others clamber ashore and IR is no exception. After the third great debate of IR theory, following Waltz's (1979) *Theory of International Politics*, what has ensued is a period of – as Duune, et al. (2013) note – "theory testing". There seems a strange discrepancy between the prominent work that is published in IR journals and the theories that dominate (or have historically dominated) IR. There seems to be some acceptance of the different theoretical paradigms in IR, the "-isms" seem widely accepted as part of the underlying logic of IR study. Theoretical hegemony no longer seems important to IR scholars, and the sometimes destructive (or at least wasteful) efforts at imposing one theoretical perspective over others has largely vanished. The pursuit of theory has been steadily replaced by the pursuit of more practical knowledge. While there seems to be evidence of this shift, it seems strange that the centres of higher learning focus so heavily on these largely ignored paradigms (Baron, 2014). I will argue that the proliferation of a wider range of theories now permeates IR scholarship, and that this proliferation – which moves past many of the assumptions of IR's "traditional approaches" – strengthens the relevancy of International Relations research.

It is clear from the history of IR that it faces a significant challenge. The field imports a wide range of insights and theories from neighbouring fields, including but not limited to: philosophy, social and political theory, history, law, economics, maths, statistics and biology. Despite this wide import, empirical evidence illustrates that very few IR scholars are able to make a significant impact on other disciplines and enter the realm of public intellectual (Duune, et al., 2013). This might be the result of the lack of more empirically focused study.

Indeed, the theory-driven approach of IR, and its concerns with theoretical paradigms are significant compared to other disciplines (Waever, 2013).

Baron (2014) makes an excellent point. In the US, scholars of IR present a view of IR defined by the various leading paradigms. However, when analysing the literature these scholars produce, they move away from such paradigms, focusing on niche, non-paradigmatic issues. This increases the potential for misunderstanding about what exactly IR is and what it studies. IR is likely misrepresented in classrooms, and this may be why IR research fails to diffuse into other fields; IR scholars themselves are confused about the research agenda. It may be that, as Allison & Zelikow (1999:404) suggest that the belief (by some scholars of the time) that the answer to the questions of "whether states go to war, join alliances, or compete could be found exclusively in system-level variables was misguided."

The relationship between individual decision-makers and institutions is a complex one. It is yet unclear how individuals form a state, or how exactly political elites influence the decision of "the state". I posit here, however, that individuals (especially leaders) significantly influence the state. Therefore, focusing on individuals as the unit of analysis may be a relatively good proxy for the behaviour of states in the international system. I also posit that these actors, though capable of rationality, are affected by psychological biases they may not be able to mitigate effectively. A move away from the traditional concepts of IR, at least in some respects, may prove to be fruitful for IR scholars, as I hope this study proves.

The focus on exogenous factors in interpreting the international world harms the study of IR. I do not think that focusing (exclusively) on system-level variables can constitute a complete theory of IR. Relationships may indeed by systematic, rational and informed by power struggles or ideas of a democratic peace. International relationships, however, are frequently personal and prejudiced, and informed by personal relationships between individuals, and indeed by the flawed decision-making processes that accompany them.

4. (Mis)Behaviour?

The language that we use matters. It is likely the case that, if we characterise certain events as mysterious, anomalous, or irrational, we disregard it too quickly. I believe that, by merely labelling decision-makers as confused, misguided or even nefarious, we lose the will to empathy and – possibly more importantly – understanding. The pursuit of knowledge is not furthered, because we do not attempt to understand that which we consider obvious. After all, we are perfectly sane, and the realm of such confused men or women does not belong to us. We reason that we could never understand it, or perhaps we give too-simple explanations, waving a hand as if to say "some people are just (insert insult)". In fact, why should we make room in our rational minds and models for the defects of these clearly delusional individuals? These other decision-makers are the anomalies, they are the outliers, and a good model cannot – should not – focus on these outliers when formulating its hypotheses and frameworks. This type of thinking is dangerous, not to mention intellectually arrogant, because we tempt ourselves to dismiss that which we don't understand under a shallow guise. In fact, what we construe as misbehaviour by other actors may simply be behaviour – that is, the cognitive functions we all possess playing themselves out in a particular way.

In the following sections, I will attempt to clearly outline how rationalist theories would interpret a specific event and then show how BIR can better account for the decisions made concerning that same event. I will systematically use the tools introduced in Table 3 above to explain some of the errors in judgement exampled by the events I discuss.

4.1 Non-compliance in International Relations: Enforcement and Preference

Compliance and non-compliance in International Relations is an interesting topic, and it provides IR scholars with a number of puzzles to solve. Interesting questions include: Do leaders attempt to benefit from false promises, why or why not? Does the fear of non-compliance of other states deter international cooperation? What role do enforcement mechanisms play in international agreements? Which preferences increase/decrease the likelihood of international cooperation?

These questions are usually answered theoretically, and there is no shortage of theories. Arguments from Realism indicate that states might default from promises in order to profit, especially strong ones. Others might argue that fears of just such defaults may harm international cooperation. Perhaps enforcing promises would cause states to think twice about defaulting. Others might argue that the cooperative nature of the human being is fundamental to the international agreements. All of these explanations are plausible. What is important is the context of the event in question, the nature of the decision, and the preferences of the decision-maker. BIR brings to the fore an explanation that fits the data more closely than theoretical predictions. Subsuming rational choice, BIR moves beyond paradigmatic analysis into the realm of behavioural psychology in order to create firm foundations for understanding strategic choice and international legal cooperation.

Hafner-Burton, et al. (2014,2015) analyse compliance in international agreement experimentally, using political elites in the United States as well as university students. Their findings indicate the importance of individual personality and preference in understanding decision-making at an international level. Their results indicate consistent theoretical validity and indicate that behavioural work in International Relations has the potential to improve our understanding of the international world, its puzzles, and even improve our predictions of the future. In an analysis of *Decision Maker Preferences for International Legal Cooperation*, Hafner-Burton et al. (2014) show that the patience and level of strategic reasoning of individuals significantly determine their preference for international cooperation, where more patient and more strategic individuals preferred larger levels of international cooperation. Very interestingly, strategic reasoning (a behavioural trait) was more significant in determining the preferred level of cooperation than was the nature of the enforcement mechanism of the agreement in question. We can extrapolate from these findings that the characteristics of relevant political elites are likely explanans for the results of international agreements. The more patient and strategic a political elite, the more likely it is that she will favour high levels of international cooperation.

Other behavioural results also help to explain the level of international cooperation. Hafner-Burton et al (2015) show that the presence of institutional enforcement lowers the likelihood that policymakers will join international agreements. This runs contrary to the expected outcome, where enforcement mechanisms have thought to lower the risk of defection by all

countries and thus increase the likelihood that states will cooperate. The evidence the authors provide for this conclusion is at the elite level, and so is unique in its sample. In the context of U.S. foreign policy, it seems policymakers are reluctant to make false promises. The desire to uphold international norms, reputational loss and fear of informal retaliation may explain this trend. These are psychological and behavioural traits, and they vary across actors. The heterogeneity of policymakers makes generalising problematic, but there is good reason to believe policymakers may have certain characteristics that distinguish them from general populations, including the ability to think strategically and to be patient (Kertzer and Tingley, 2018:10). This is highly relevant, as the characteristics of the relevant decision-makers likely influence compliance in IR to a larger extent than the enforcement structure of a particular agreement. Flexibility in creating international agreements seems to be important in creating achievable and demanding goals that states will actually opt-in for.

Behavioural studies in IR provide a more psychological-behavioural foundation from which to understand IR phenomena. A behavioural foundation certainly seems to explain the issue of non-compliance in IR well, both from a theoretical and empirical perspective, where experimental findings are externally valid. What matters most in assessing the likelihood of compliance or non-compliance in an agreement seems to be the reluctance of political elites to make false promises. In particular, individuals who are patient (have lower future discount rates) are more sensitive to the prospect of not honouring commitments. In the presence of many patient policymakers, provisions allowing non-compliance or other flexible alternatives may actually increase cooperation (Miller at al., 2015).

Trade agreements are likely to be more widely accepted if they consider the individual preferences of the policymakers in question. A system and catalogue that measures and stores these preferences might be an extremely useful tool for speeding up negotiations in the international realm (Hafner-Burton et al., 2015).

A brief summary: there exists evidence that, at the elite level, as the risk of future defection rises, the likelihood that policymakers will join a given international agreement decreases. An additional finding runs counter-intuitive to the logic of many theories of international institutions. The existence of a formal enforcement mechanism does not seem to explain this aversion to cooperation (Hafner-Burton, et al., 2017).

How might Realism plausibly approach this puzzle? To begin with, it is important to note that making false promises presents elites with more than a few problems. Risk and uncertainty are common to political elites, and scholars typically conceptualise these risks by envisaging these environments as strategic games of incomplete information, where information about the intentions and payoffs of other players are not clear (Hafner-Burton, et al., 2017). Of course, given these risks, there is a cost associated with non-compliance. False commitments to international agreements are relatively common in IR. Realists would likely note that the costs of non-compliance are frequently less than the gains. The Russian Federation, under Communism, frequently paid lip service to human rights commitments that were supposed to be legally binding while flagrantly reneging on these commitments (United Nations, 2016). Where human rights institutions are weak or non-existent, or the law is weak and practically unenforceable or where the international community and the norms of the international community would prevent significantly punitive reaction, it may be beneficial to renege on international promises. Another example includes the fact that the US signed the Kyoto Protocol despite clear evidence that they could not, and clearly did not intend to, honour their commitment to cut their emissions. This is reminiscent of the US position on the more recent Paris climate accord (Hovi, et al., 2012). Countries like Bangladesh have also committed to promises they did not fulfil, seemingly in order "protect Bangladeshi workers in exchange for receiving Western trade privileges" (Hafner-Burton, et al., 2017).

Realists could all reasonably point to these examples in order to conclude that a state may make a false promise in order to further its own interest. There are many instances where such behaviour would be deemed strategically logical. Given this, the following assertion would be that the risks of making false promises should not heavily influence the decision of a state to join international agreements. Additionally, a formal enforcement mechanism would disincentivise joining international agreements, given a risk of non-compliance. Therefore, the willingness of a state to join an international agreement should be inversely related to the strength of the enforcement mechanism, so that strong mechanisms more significantly reduce the willingness to comply.

Simmons (1998) discusses compliance in IR from a realist perspective. Power, not law, is the primary variable in interstate relations for a Realist. The author goes as far as to state that "most realists - theoreticians and practitioners - tend to be highly sceptical that treaties or

formal agreements influence state actions in any important way" (Simmons, 1998:5). According to this perspective, compliance and the observation of international law is the result of converging interests or prevailing power structures. Realists thus view promises in IR as tenuous and breakable.

The Realist assessment would not come to the conclusion that has been empirically demonstrated by Hafner-Burton, et al. in their 2015 study. Realists i) expect false commitments to made relatively often, if it benefits the state and ii) expect that strong, credible enforcement mechanisms would increase the cost of reneging, to the point of making it a poor strategic decision, and lower the odds of cooperation. In both cases, the Realist view does not satisfactorily explain the result of the authors.

What about the Liberal view? A liberal thinker would likely point to the benefits of international cooperation. They might argue that international agreements mutually benefit the involved states. A trade agreement might allow sectors to specialise along lines of comparative advantage, for example. A liberal thinker might indeed predict that international norms guide behaviour, but would likely look to structural variables to predict what form the behaviour would take. For example, a liberal argument might be that the international norm of keeping promises is one that benefits all states and thus if states cannot fulfil a promise, they are hesitant to commit to an agreement. Simmons (1998), argues that democracies are more predisposed (than non-democracies) to comply with international agreements of all kinds. Two lines of reasoning support this argument. First, democratic states respect and conform with legal processes and international institutional processes, and display a willingness to depend upon the rule of law of external as well as internal affairs. Of course, this depends on the assumption – made by many liberal scholars – that norms of limited government, a great respect for legal processes and emphasis on constitutional constraints are exported into the international arena. It is plausible, therefore, to believe that liberal theorists would argue that those governments with strong constitutional traditions (such as the United States), are more likely than other governments to accept a measure of law-based constraints on international behaviour.

How does the liberal perspective stack up to the empirical finding of the study in question? While a liberal theorist will likely predict that, given the respect agents in IR have for lawful

processes and institutions, under increasing prospects of defection states will become less willing to commit to agreements. Reputation, in liberal theory, is important. In a morally concerned world, where cooperation is tantamount to prosperity, honesty is vital. In an iterated game, where cooperation is mutually beneficial and war is catastrophic, it makes sense to have a reputation for keeping your word. Liberalism, however, fails to account for the second finding of the study – namely that credible enforcement mechanisms do not explain any existing aversion to cooperation in political elites. For the same reasons that liberal theory accounts for the first finding of the study, it fails to explain the second. If states of political elites are particularly respectful of international institutions and international law, formal enforcement mechanisms should significantly reduce the willingness of political elites to engage in related agreements. However, Hafner-Burton, et al. (2015) show empirically that formal enforcement mechanisms make it only three percent less likely that a political elite will opt out of an international agreement. In a word, formal enforcement and a great respect for international laws and regulation do not seem to explain an aversion to committing to international agreements.

The constructivist approach would likely emphasise the existence of shared norms and beliefs (international society) as a crucial ingredient in the function of international law. Simmons (1998:14) notes that the constructivist view argues that the "primary function of international law is helping to mobilize compliance with the rules of what is termed international society." Constructivist theory delves into the inter-subjective interpretations of non-compliance, and notes that the perceived legitimacy of a given international legal rule increases compliance through imposing a greater sense of obligation on the actors in question. I would argue that the constructivist view would fail to account for the issue in question for the same reason as the liberal argument. A respect for legitimate law seems not to influence compliance. Admittedly, the (non)compliance with inter-state agreements is a fairly narrow field, and it is difficult to propose here a convincing account of expected behaviour in this area stemming from the critical theories of IR. These theories are concerned with issues larger than this rather specific example, and therefore I will not discuss these critical theories in this section.

Does Behavioural International Relations give a more convincing explanation for the somewhat surprising behaviour we have been discussing? Hafner-Burton, et al. (2015)

investigate the traits of their participants, one of which is their patience. They (tentatively) posit that the aversion to cooperation – even given the lack of credible enforcement mechanisms – is linked to the individual perceptions of political elites. In particular, the authors cite the discount rates as the pivotal factor here. In Table 3 on page 39, I outline a particular cognitive bias called 'hyperbolic time discounting', which is the tendency for individuals to discount the future at a growing rate. Hyperbolic time discounting is a particular cognitive bias based on time preferences of individuals. Whether or not political elites have hyperbolic time preferences is not important, but Hafner-Burton, et al. show that time preferences in general are important in this context. Namely, they propose that patient individuals are more sensitive to reneging on their promises and are thus averse to cooperating in international agreements, no matter what enforcement mechanisms are in place.

By way of summary then, elite politicians are heterogenous in their analysis of non-compliance risk. Those who have longer-time-horizons are less likely to comply, while those with shorter-time-preferences are more likely to comply, since they are less worried about the costs of making false promises. The weight that decision-makers assign to the future may be more important than the structure of international agreements in considering the cooperation of states in agreements. Traditional International Relations theory fails to provide convincing accounts for this puzzle. Perhaps these theories are not meant to solve puzzles such as this, it is a narrow area, after all. Well, whether these traditional lenses themselves claim to solve such puzzle is irrelevant. Either they do aim to solve such puzzles, and they fall short, or they do not, and we need a more precise theory to deal with them. BIR proves useful in this case, either way.

4.2 The Centrality of Beliefs in Foreign Policymaking: The Case of Tony Blair

Do we have beliefs? Or do beliefs have us? I made this point earlier in my analysis. It is one that is fundamental to behavioural studies, because the answer determines the direction in which causality in decision-making runs. If we have beliefs, and we change them rationally in response to experience, we can behave increasingly rationally. If, however, we allow confirmation bias to convince us that the world is a certain way, and we are not critical, we may become ideologically possessed. That is, it is possible to be blind to experiences which

should cause us to change our beliefs, assuming we wanted to behave rationally. Even political elites, maybe especially political elites, are susceptible to ideological possession (Tetlock, 2005).

Nhandara (2015) embarks on a case study of Tony Blair's operational code and shows that Blair did not manifest statistically significant changes in his beliefs over the course of his time as British prime minister. What explains this? Certainly, it seems highly anomalous given traditional expectations of rationality and strategic thinking by political elites. Robinson (2011) also finds that there is little evidence for experiential learning by political elites, save for major shocks and abundant contrary information. Even this case however, effects are not always in the expected direction. Nhandara (2015) examines the operational code of Tony Blair during the period from 1997-2007 when he was prime minister. A hypothesis is laid out by the author, one I believe would be held by any theoretician in International Relations subscribing to a 'dominant IR paradigm'. The hypothesis was that, in the wake of the September 11 attacks, Blair would shift his philosophical beliefs and show signs of significant experiential learning. This seems not to have been the case, despite exogenous shocks which – in structuralist literature – should have resulted in a measurable change in foreign policy behaviour (caused by shifting beliefs). The author divides events into three time periods, "Post Kosovo – Pre Iraq, Pre-9/11 – Post 9/11 and Pre EU – Post EU" and analyses operational indices in these periods to determine any changes in beliefs. The author concludes that these events provided no statistically significant changes in Blair's beliefs.

Dyson (2006:289) argues that the British involvement in Iraq "has been characterised as 'Tony Blair's War" and that the specific personality profile, leadership style and operational code of the prime minister was instrumental in the British involvement in the middle east. Indeed, Dyson provides plausible evidence that Blair's personality traits significantly shaped the British foreign policy with regards to Iraq, and reaffirms the research that emphasises the importance of individual-level features in IR theory and foreign policy analysis, in addition to positing a compelling explanation of the British involvement in Iraq.

These anomalies are not predicted by most IR theories, which expect experiential learning to take place, and expect that state interests are the primary factors in a decision to go to war (not individual personality, as may be more influential than we've previously thought). What

explains some of the reluctance of leaders to change their beliefs? From a BIR, and psychologically informed perspective, there are likely explanations. Tetlock (2017) notes the centrality of reputational considerations in adjusting belief systems. Reputational loss for an expert is a significant threat, and (for obvious and mundane reasons) reluctance to change beliefs is in part a stubborn attempt to maintain legitimacy. This is a most human problem, rooted often in ego rather than rationality. In addition to these effects, the Sunk-Cost bias that is widely known and outlined in Table 3, causes individuals – even leaders – to stick to their course of action because of their previously invested efforts and material resources, even if circumstances make doing so unprofitable. Another explanation is that those politicians who have a strong belief in the ability of a core variable to explain almost all phenomena receive wide attention. Intellectual aggressiveness and confident predictions make more interesting television than, say, a tentative political forecaster who acknowledges the many complicated variables involved in an event, and who is a strong and diligent statistician. It is conceivable, though regrettable, that political elites who attempt to expand the explanatory power of their one key variable are more interesting, more sensational, and more represented in mainstream media and therefore rewarded monetarily – at least in the short term. Learning is painful, to be blunt, and self-deception is easier than we might expect. Poor methodological, critical and cognitive methods all contribute to a failure to update beliefs, and not all of these failings are rooted in arrogance. Tetlock (2017) certainly underpins the importance of scientific and systematic rational thought, as well as a large measure of epistemic (rather than political) motivation and intellectual humility when studying, understanding and predicting phenomena in IR.

Leaders appear to, as Tetlock (2005) notes, dissociate themselves with their own belief systems, which are often unconscious to the individuals. Nhandara (2015:76) makes the claim that “leaders do not always respond in a rational manner towards foreign policy issues”. This might well be because their responses are processed (as I have suggested) by a system of beliefs and perceptions they themselves are unaware of.

Referring back to Table 3 on page 39, there are a number of interacting cognitive biases in the specific case of Tony Blair. In the first place, the personality of Blair shows a tendency towards the Overconfidence bias, evidenced by his high belief in his ability to control events (Dyson, 2006). Blair also has a tendency for low conceptual complexity, a proxy for both the

Framing Effect bias and the Understanding bias (which is the tendency for us to overlook pertinent but complex information in favour of simple, easily interpretable date). Blair's personality is predisposed to frame 'others', for example, Hussein, not only as an 'enemy' but further as an evil dictator, as deceitful and other negative frames. He is also predisposed not to look for further explanations for Hussein's behaviour, since his frame of the situation provides an emotionally compelling and coherent (if not entirely accurate) narrative.

In addition to the role that Tony Blair likely played in the Iraq invasion, he seemed not to undergo experiential shifts in his beliefs. This finding is not unique to Blair, and it is somewhat anomalous (Robinson, 2011). After all, rational actors should be able to shift their beliefs in the face of compelling evidence. One compelling explanation to this puzzle is evidenced above. The personality of individual leaders, overconfidence for example and the costs to reputation for a belief-changing politician (and the Sunk-Cost bias, which causes us to overvalue courses of action we have invested in) cause individuals to maintain their beliefs and actions even when it would be strategically rational to do so, both from a state and individual perspective.

4.3 The Invasion of Iraq, bizarreness and bias

Hindsight is a perfect science. Of course, we must be wary of framing the past as inevitable, as though it should have been obvious. It rarely is. Yet, by all accounts, the US led, costly and poorly planned invasion and occupation of Iraq was bizarre. It birthed conspiracy theories, was hotly debated and constitutes one of the most controversial issues in international politics in recent history.

A most interesting perspective on the Iraq war is given by Harvey (2011) in his book "*Explaining the Iraq War: Counterfactual Theory, Logic and Evidence".* Of course, the war initiated by the Bush administration in 2003 was divisive on a number of fronts. The academic discourse likewise remains without a clean consensus on the causes and nature of the episode. Particularly interesting in Harvey's analysis is that it gives both an account of the invasion of Iraq and a contribution to political science more generally. As I am giving an account of the importance of the notion of cognitive bias and the methodological style of BIR, it is vital to note the important inclusion of empirical data in political analysis as a basis for theoretical justification.

What makes Harvey's analysis of the Iraqi war interesting? More importantly, how is it relevant in aiding the argument that Behavioural IR is better at explaining the Iraq war than other theories? Let us first present the predominantly accepted narrative of the war on Iraq that began in 2003. Harvey argues that the war "a product of the political biases, misguided priorities, intentional deceptions and grand strategies of President George W. Bush and prominent neoconservatives, 'unilateralists' and 'Vulcans' on his national security" Harvey (2011:1) Neoconservatives in this definition represent a political ideology characterised by an emphasis on free market capitalism as well as interventionist foreign policy. The narrative following on from this basis presumes that a neoconservative dominated Bush administration was a necessary condition for the war to occur, and that a Gore administration would thus represent a necessary condition for peace in Iraq. As you may have guessed, Harvey does not endorse this view. The author makes use of numerous analytical categories including: personality characteristics, political outlook, the beliefs and perceptions of Gore's political analysts and advisors, bureaucratic elements, public opinion and even Saddam Hussein's own (mis)calculations and manoeuvrings to make a strong counterfactual case. That is, the case that a Gore administration would likely have gone to war with Iraq too.

It is not my aim here to present a thorough breakdown of Harvey's entire book. What is crucial is that there exists an empirically justified counterfactual account that is found to be relatively compelling by experts in the field (Dawisha, et al., 2013). The key take-away here is that strange strategic behaviour is explained by a BIR approach in a way that makes it understandable and – potentially – predictable. In particular, it opens up the possibility of greater nuance in our political analysis.

The Iraqi war may not solely have been the result of the Bush administration and the neoconservatives. Misinformation and strategically miscalculated behaviour (for example, Hussein's refusal to demonstrate Iraq held no Weapons of Mass Destruction [WMD]), systemic factors, personality traits and cognitive biases and strongly held beliefs all contributed to the Iraq war. In a word, individual, internal factors as well as external, systematic variables are both important in IR, and Harvey's meticulous presentation of strong empirical evidence gives us a compelling reason to take a leaf out of his book, in order that we might strengthen our academic toolkit. What are some of the characteristic cognitive

biases and individual-level factors that BIR would argue played an important role in the invasion of Iraq? Let's begin with the case of Saddam Hussein. Post (1991), examined the personality of Hussein far before the Iraq war, which provides interesting insights. Hussein was found, by Post, to hold a narrow and distorted worldview. His actions were ideologically rationalised and he had powerful messianic goals. He had a "strong paranoid orientation" and saw himself "as surrounded by enemies" (Post, 1991:284). The issue of foreign domination was crucial for Hussein. He saw his refusal to bow to the West and his wish to expel foreign influences as part of a powerful self-held myth. These personality traits are not conducive to negotiation and transparency. It is highly likely he saw capitulation to the West's requests for assurances as weak, which conflicted with his view of himself. Perhaps it is not surprising, therefore, when he did not credibly assure the US that Iraq was not going to develop WMD in the future. Hussein's ideologically driven thinking was certainly prone to biases. Among them was a powerful Framing Bias, that – exacerbated by his paranoia – would allow him to see nothing but traitors and schemers in US politicians. His strong emotion almost certainly played a role in his perception of all information. The Emotional Bias likely aggravated his ideological beliefs, and Confirmation Bias caused him to interpret all available information along the established narrative: The West seek to destroy Iraq, and to destroy me. It is easy to dismiss Hussein's behaviour, but he was not psychotic. There is a place in every human being for powerful narratives, self-deception and dangerous biases, and we should be wary of that fact.

Hafner-Burton, et al. (2017) also note that the invasion of Iraq and the consequent failure of Saddam Hussein to prevent it, can explored with traditional theory. Rationalist tools would argue (credibly) that there existed asymmetries in information, among other issues. But these arguments leave much to be desired. For example, as Hafner-Burton et al. (2017) state, "the Clinton and Bush administrations did not differ substantially in their information about Iraq. But Bush administration officials—and the president himself—did hold beliefs that differed substantially from those of their predecessors, and those beliefs had profound effects."

Bargaining theory certainly had a role to play in the Iraqi invasion, but so did beliefs, decision-making and biases. Colin Powell, American statesman who served in under George W. Bush stated, in 2003, that "the liberation of Iraq is a great victory for freedom. It has freed the

international community from the threat posed by the potentially catastrophic combination of a rogue regime, weapons of mass destruction and terrorists. And it has freed the Iraqi people from a vicious oppressor" (The Guardian, 2003:1). Clearly, Hussein failed to communicate credibly that he had dismantled his weapons of mass destruction. In what might have been a comedy of cognitive errors – if not for the associated suffering of the Iraqi people – Bush and his administration thought that they would be welcomed and democracy would flourish because (in their estimation) it was just. This indicates the Wishful Thinking Bias clearly. In addition, reasoning by wrong analogy severely hampered the reconstruction of Iraq after the invasion (David & Prémont, 2015).

The US involvement in the Iraq war was extremely costly on a number of fronts, and the rationalist accounts of the "why" of the invasion are incomplete. As Checkel (1998) notes, rationalist IR theory is criticised, not because of what it says, but rather what it ignores. In the case of Iraq, there is much that is ignored by rationalist theory that is an important part of the puzzle. I will, in the following summary, add into the analysis the insights that Behavioural International Relations – through the study of individual belief and bias – adds to the Iraq war narrative to create a more compelling, and complete explanation.

A brief account of the rationalist account of the Iraq war here follows, derived from the work of Lake (2011). Thought there were outcomes short of war that would have left the US and Iraq better off, there are strategic contexts that aid in our understanding of its occurrence. I) Saddam Hussein couldn't credibly communicate to the Bush administration that Iraq would not developing WMD. II) Information asymmetries, where both Iraq and the US overestimated their abilities to impose costs on each other, and drastic underestimations of the costs of war and post-war operations caused the two parties to engage in a failed and costly conflict for both sides.

Rationalist IR theory does not account well for the self-delusions that were present in the Iraq decision. There was an unsuccessful and incomplete process of information gathering and considering alternatives, which is expected in bargaining theory. Lake (2011) notes that "decision-making in the Bush administration and Saddam's regime, as far as we can now tell, did not fit… [rationalist models of the Iraq war]." On the U.S. side the Bush administration illustrated a typical Overconfidence Bias (see Table 3 on page 39). The Bush administration

underestimated its own costs of war, and overconfidently believed that post-war reconstruction in Iraq would be relatively smooth, erroneously assuming democracy was favoured by the citizens of Iraq. In addition, Confirmation Bias also reared its head, and Lake (2011) notes that the administration suppressed the estimates of the war's costs given by outside observers and critics. Together with the Bush administrations ideologically rigid views and the dominant US narratives as the champions of freedom, these biases lead to a failure by the US to plan realistically for post-war reconstruction in Iraq.

Lake (2011:52) sums up my own argument very neatly, and is worth noting at length:

"It would be inappropriate to reject bargaining theory on the basis of a single possibly extreme case. Nonetheless, the Iraq War brings the effects of cognitive biases and human fallibility into sharp relief. In addition to extending theory to include a post-war bargaining phase, moving to n-player models of signalling, and incorporating the role of special interests, the Iraq War suggests the need for a behavioural theory of war that integrates human decision-making biases into the strategic interactions that, through bargaining failures, produce war. The first behavioural revolution in international relations led away from a focus on what states should do to a new emphasis on what they actually do. It is, perhaps, time for a second behavioural revolution in international relations where scholars focus on how individuals and, as collectives, states actually think."

Constructivists might understand the Iraq war as a means of constructing the identity of the US state (e.g. as the 'good guy' and 'saviour' in the international arena, the 'protector' of the world, who swoops in to prevent mass destruction). Their arguments should be noted and connected with the Behavioural IR to understand the psychological motivations of individuals in what seem like very risky (costly) decisions, though the project of doing so is not within the scope of my work.

4.4 The Forecasting Inaccuracy of Political Experts

Phillip Tetlock (2005) published a book entitled *Expert Political Judgement: How Good Is It? How Can We Know?* In it, he discusses the results of a twenty-year study on expert political predictions. The study incorporated 284 people, all of whose professional capacity included the "commenting or offering advice on political and economic trends."

The database he evaluated consisted of over eighty thousand forecasts. The huge amount of data is one reason to take the findings seriously, in addition to the quality of the method and study. The predictions of the experts were measured against alternative predictions, including simple statistical models and non-experts. The experts did not outperform non-experts at forecasting, and neither experts nor non-experts outperformed simple rules and models. The result is not a completely novel one either (Tschoegl & Armstrong, 2007).

Is this an issue? I think a reasonable analysis would conclude that it is. Experts, who are often theoretically driven (especially in IR), fail to predict outcomes more accurately than – essentially – random chance. Surely, this observation indicates the relative impotency of a particular theory. Indeed, Tetlock (2005) makes this point. Those forecasters that were relatively more ideologically driven in their predictions performed worse than those who used multiple and varied sources of information. I would argue that the polarisation of theoretical perspectives in IR, i.e. the split between the "-isms", encourages dogmatic views of the world, severely harms the predictive – and by extension, explanatory – power of these theories.

There was also evidence that those participants with the most expertise were less effective at using new information, and thus failed to update their expectations adequately. Ideological possession, in a word, is not useful. The truth is more nuanced than the dominant paradigms assume (especially Realism, Liberalism and all their derivative forms). They may still be useful however, but not alone. Incorporating them and creating a more nuanced view of the world promises to increase the relevancy of these paradigms.

How can a psychologically informed approach, such as Behavioural IR, provide support to traditional IR theory in order to improve forecasting? Mellers, et al. (2015) provide insight. Three key categories of variables seem to increase forecasting ability in individuals. These are i) Dispositional, i.e. cognitive aptitude, political knowledge and openness; ii) Situational, i.e. levels of training in probabilistic reasoning and collaboration with peers and iii) Behavioural, i.e. frequency of belief-updating and the time spent by an individual before making a given forecast.

Mellers, et al. (2015) illustrate that, when making predictions about the political future, engagement is central to the success of these predictions. That is, the willingness for

predictors to reform decisions, to update their beliefs, benefits their predictive ability greatly. Where participants view learning as a skill, were likelier to perform well in prediction, than those who assumed that their learning was complete at the stage that they gave a prediction. In addition, the time that a participant spent viewing, gathering evidence and deliberating before making a forecast positively correlated with their accuracy. In a word, 'doing your homework' will likely lead to better predictive results than relying on snap judgements derived from intuition.

How do we relate this back to our example? Political forecasters may not predict medium-long run futures very well at all. It is quite plausible that a related cause for this is that ideological stubbornness – refusing to update one's beliefs and re-evaluate one's perception – does not lend itself to learning. I argue that ideological stubbornness is one cause (among many) of poor forecasting. Traditional IR theories are not well-updated. Of course, there are constant theoretical derivatives of IR theory, and these should be encouraged. However, there is also the tendency for IR scholars and commentators to give exaggerated claims and dress them up as gospel truth. Hopefully, psychological behavioural work can continue to show that it is a more pragmatic approach to remain open to criticism and learning from outside. In this way, we are more likely to move towards better explanatory and predictive models of the international political world. Behavioural IR aims to play a role in the updating of the IR belief system, and in this way continue to move it forward.

4.5 Hyperbolic Discounting and Climate Change Negotiations

Traditional theoretical modelling and thinking fails to account for the degree to which decision-makers seem to discount the future. Models have always discounted the future – that is, people value gains/losses in the future less than they do equal gains/losses in the present – but these models fail to explain certain phenomena, like climate change. Karp (2004) shows that "Hyperbolic discounting is a plausible description of how people think about trading-off costs and benefits in the distant future."

The concept of hyperbolic discounting is an insight that the behavioural literature has played a significant role in uncovering. Hyperbolic discounting simply refers to the tendency for decision-makers to value the future at an exponentially lower rate than recent events. When

plotted on a graph, if one puts discount factor on the y-axis and time from present on the x-axis, the graph represents a hyperbola. Thus, people tend to choose smaller reward occurring sooner over larger rewards occurring later.

The concept of hyperbolic discounting is counter-intuitive to rationalist theory. Measuring losses and gains in this way is not rational, it is certainly inefficient, and harms society in many ways. Quite apart from the inefficiency of such time preferences, I would argue that the phenomenon is also a moral issue. How? Hafner-Burton, et al. (2017) point out that "countries have agreed on bold goals, such as stopping warming at 1.5 or 2 degrees Celsius above pre-industrial levels, yet they pursue policies likely to cause perhaps double that warming by 2100 and even more in the years beyond." Global climate change could seriously harm (to put it lightly) the well-being of future generations. We have a moral obligation to alter our discount preferences to look after the well-being of the future of humanity, no matter that they do not yet exist.

Another explanation for understanding the relative lack of urgency over climate change, given the momentous potential losses, is given by Peterson (2017). The author concludes that "individuals with high system justification discounted climate change more than those with low system justification." Thus, an economic and social system which relies heavily on fossil fuels (for example) is likely to spawn decision-makers who are less willing to change the status quo than those systems which place a large degree of importance on renewable energy. An implication of this is that, where there exists a destructive cycle in, say, the US – which contributes 45% to world pollution – there will probably exist reinforcing cycles where countries systematically endorse renewable energy.

This is, perhaps, nothing more than an elaboration on the ever-present Status-Quo Bias (see Table 3 on page 39). Those who are accustomed to a certain way of life, a certain energy infrastructure or even certain ideas of climate change are internally resistant to change. This bias is itself a derivation of the loss aversion bias. This is clear when one considers Donald Trump's views of climate change.

Once again, only Constructivism might be able to give an account of this behaviour that is satisfactory. Constructivists, of course, could point to the constructed world of Trump, in which his beliefs are certainly constructed, since they do not conform to widely available and

well-known evidence. Additionally, Trump's wish to be a strongman is likely inhibiting his ability to public dissent against his own past views, since this would make harm his narrative as a confident and unapologetic leader. Liberalism and Realism would certainly struggle to give good accounts of Trump's bizarre position on climate change. Realism might argue that Trump believes that supporting traditional energy will maintain the economic power of the United States. This self-interest is predicted by Realism. However, the U.S. is at risk of losing its position as a leader in climate change, and is losing credibility as a rational and cooperative actor. Beyond that, if we accept the views of over ninety-percent of climate change scientists, clean energy will become increasingly important in the future. It may be much costlier in the long-run to cut research funding for clean technology. Liberal thinkers are certainly perplexed by Trump. His refusal to conform to international norms, his refusal to update his own beliefs, and his US-centric view of international politics all violate models of a rational agent.

Looking past the obvious Confirmation Bias and reluctance to update his beliefs given evidence, there are other behavioural explanations for the Trump's administration attitude on climate change. Zhang, et al. (2017:1) note that "Fossil fuel industries hold powerful political clout over the Trump Administration and the Republican Party" and that that there are complex personal financial incentives in play. Further, Trump believes that the Paris Agreement weakens U.S. sovereignty, the competitive edge of the Unites States' economy and harms employment in "dirty" energy industries. It is quite possible to frame this as an issue of the Status-Quo bias. It is certain that meeting the Paris Climate Agreement will result in the loss of jobs in traditional energy sectors. These sure losses weigh heavily on the Trump administration, despite the opportunity – still hypothetical and therefore risky – to grow the renewable energy sector. Even if, on balance, more jobs could be created and the economy bolstered by investing in renewable energy, the loss and risk aversion of Trump, exacerbated by his erroneous beliefs on climate change, has led the US to halt (and even reverse) their position on climate change.

4.6 Traditional Theory and Behavioural International Relations: Possible Responses

So far, I have given a somewhat unbalanced account of the positions of the 'traditional IR paradigms' and Behavioural International Relations respectively. Of course, this is part of my argument and thus necessary for the point to be made. However, it is important to note the possible responses that traditional theorists might have in response to these comments.

Firstly, I have asserted the importance of analysis at the level of the individual, and the crucial role that individual heterogeneity has on the manner in which we make predictions and retroactively make sense of phenomena. In order to understand the decisions of an agent, we need to assess his or her preferences and beliefs. This is an area that behavioural work has expanded into. However, it is by no means entirely original. The classical distinction between 'hawks' and 'doves' (Colaresi, 2004: 555) differ in their preferences for conflict. Hafner-Burton et al. (2017:7) note that "Many scholars have modelled the preferences of agents who make foreign policy decisions, including bureaucracies, leaders, or even the proverbial median voter. The theory of audience costs provides a well-known example of this variety of theorizing." Thus, traditional IR theory does have branches of research which focus on the individual, and further – they develop models which are characterised by varying preferences. Thus, the claim that traditional IR theory deals unsatisfactorily with certain phenomena is not an indictment against the foundations of IR theory, but merely the extent and nature of its ability to update preferences in line with new research. This line of reasoning leads us to conclude that IR theory merely needs updating, not reinventing. Is a new field of IR really necessary to account for this purpose?

The second response that traditional theorists could justifiably make would progress along the following lines. I have noted that the conditions of the traditional, formal, model of expected utility assumes full information, and that this assumption is highly problematic – especially in the arena of international relations. However, one could point out that there exist game theoretic models that are constructed under the assumption of incomplete information. Indeed, "Games of incomplete information…assume only that players have common knowledge over the structure of the game and beliefs about how parameter values are distributed" (Hafner-Burton et al., 2017:7). In International Relations, signalling games for example, have been constructed which indicate the uncertainty that each party

experiences, as well as the costs of signalling and difficulties in determining the true intentions of other parties. One may, reasonably, ask why it is necessary to dismiss models of behaviour when underlying assumptions might be adjusted. These responses are worth noting, but fail to note the extent to which assumptions have to be transformed in order to approximate reality. The change is so fundamental that it may actually be easier to start from first principals, and build a theory from there.

In relation to this approach, BIR comes under some criticism. While there is much to be said for the deductive validity of traditional IR theories, and their coherent set of assumptions, inferences and observations, there is a comparative lack of unified theory in the area of Behaviouralism. This a problem I have attempted to deal with in discussions above, but it is a significant problem. It will need to become increasingly clear what BIR observes, measures, assumes and infers. The pillars for the theory are becoming clearer, thanks – in no small part – to the texts analysed in the previous chapter. However, there is still a need to develop consensus on the approaches and issues that BIR concerns itself with, in order to develop something like a complete theory of behaviour.

4.7 A Short Comment on the Utility of BIR

Beyond illustrating the exciting possibility of expanding our International Relations methodological and theoretical tools, the above analysis indicates that there is significant utility in uniting different schools of thought, theoretical views and methodological practices. The often-arbitrary divisions between subject matter makes it possible to focus on specific issues, and that is useful, but there is no reason that we should stifle the useful interplay between – especially but not limited to – philosophy, political sciences, economics, psychology and biology. The issues that are central to the human condition, and of great interest to scholars, are often best examined with a multivariate analysis and a multidisciplinary thinking hat. It is my hope that this humble work contributes to that argument, and that we will see fruitful debate across fields in order to drive our accumulation of knowledge forward, so that we may more effectively solve issues such as climate change, war and violence, famine and economic uncertainty.

5. Conclusion

This chapter asked: Are strange events in International Relations are the result of misbehaviour or of misunderstanding? The limits of dominant IR paradigms were examined. First, it was illustrated that these paradigms still play a role in educating students of International Relations, but that this may be harming rather than helping the field. The chapter highlighted the main tenets of these paradigms, and showed that these tenets are not capable of satisfactorily explaining some events in IR.

These events included (non)compliance in IR, beliefs in policymaking, the invasion of Iraq, the forecasting inaccuracy of political experts and climate change negotiations. The chapter argued that incorporating ideas belonging to the toolkit of Behavioural International Relations and related literature improves our ability to understand and explain these events. Ultimately, the chapter laid the foundation for considering the possibility that what IR scholars, historically, have considered misbehaviour on the part of IR agents, is actually misunderstanding. Cognitive biases play a role in correcting this misunderstanding, so that we can more accurately specify what we should expect from human decision-makers.

This chapter presented some potential responses that traditional theory might give BIR in response to its criticisms. BIR is not a complete or unified theory in the sense of the more traditional approaches, but is moving in that direction. It also provides a better approach, comparatively, and will be useful for scholars and policymakers in the future – to the extent that it provides stronger micro foundations for a theory of human behaviour.

Chapter 5

Core Arguments and Future Research

1. Introduction

In this chapter I will outline the core arguments of the paper. I will attempt to analyse these arguments and determine to what extent I have answered my research question: In light of "anomalous" behaviour by actors in IR, often reduced to irrationality by traditional IR theory, might BIR – which focuses on cognitive bias as a core variable in its analysis – better explain such behaviour?

I laid the foundation for mutual understanding by creating (or at least explicating) a conceptual and theoretical framework with which to understand the concept of cognitive bias. I outlined the explanatory limits of dominant IR paradigms in relation to certain puzzling phenomena. I also showed that IR studies suffer from being trapped in an echo chamber, where IR scholars import but rarely export their ideas into other fields. I illustrated the limits of rational choice in decision-makers, and expounded concrete events that do not line up neatly with dominant IR paradigms. I then used an integrative review to show the insights, process and applications of BIR (and the closely related concept of cognitive bias). Further, I used explanatory examples, garnered from research in IR, to show the evidence suggesting that the field of BIR does indeed provide compelling explanations for the bizarre events I identified previously. In this chapter, I will defend my assertions against counterpoints. I will determine to what extent I have answered my research question with the study and suggest which areas require future research in order to strengthen and further these findings.

2. Core Arguments, Justification and Responses

2.1 On Cognitive Bias

In chapter 2, I outlined a conceptual framework for cognitive bias. I did this in order that i) it would be clear that the concept of cognitive bias is crucial in Behavioural IR and central

to its study and ii) we might have some common ground from which to understand the concept and think about it further. In my examination of the BIR literature I encountered numerous frameworks for the concept of cognitive bias.

As far as I can tell, two major taxonomies exist in the behavioural literature for classifying and attempting to solve cognitive biases. Montibeller & von Winterfeldt (2015), following Arkes (1991), propose a taxonomy for classifying biases according to psychological origin. This taxonomy breaks down biases into three categories: Strategy-based (SB) errors, Association-based (AB) errors and Psychophysically-based (PB) errors. SB type errors appear to be the easiest to correct for, while AB and PB errors are more elusive in detection and more difficult to solve. Hafner-Burton et al. (2017) use a system essentially derived from Kahneman & Tversky (1979). This taxonomy categorises cognitive bias inasmuch as they differ from rational-choice expectations. The core classifications are i) Nonstandard Preferences, ii) Nonstandard Beliefs and iii) Nonstandard Decision-Making Procedures, where "nonstandard" refers to behaviour that is not expected by rational choice theory.

Why have I deviated from these already established categories of cognitive bias in my analysis? Cynics might make the point that I am further dividing the topic into unnecessarily differentiated taxonomies that serves to widen rifts already present and harming the creation of the common language behavioural scholars are attempting to build to enrich the research of BIR. I seek, in fact to do the opposite. I do not think that the existing taxonomies are particularly intuitive, or that they serve scholars well in presenting an accessible snapshot to policymakers and other relevant non-academics. Since the study of IR is inextricably linked to those outside the field of academia, and there is a great need for dialogue with these individuals, and indeed groups, I think it important to create a conceptual space that is intuitive for those unfamiliar with the scholarly literature. I have used the existing taxonomies in IR to create one that I believe to be more intuitive. Since we usually think in linear time, it is quite easy to understand the movement of information from the external world through to our cognitive faculties in various discreet stages. Of course, this is not a perfect picture, since – as I mentioned – cognitive processing is not conducted discreetly, and different stages of the process occur simultaneously as well as create feedback loops. Still, classifying these processes as

discreet does not greatly undermine their causes and effects, and is fairly easy to intuit. While it is possible to debate the extent to which I have created a synthesis of taxonomies in an intuitive way, that has been one of my primary goals.

2.2 On the Limits of Dominant IR Paradigms

There is significant frustration, at least among some scholars, with the Great Debates of IR and the dominant Paradigms of IR. Like everything in the field, however, even the necessity of the so-called "-isms" is contested. Of course, classifying sets of ideas broadly does provide an easy reference point – one which we are familiar with and can clearly identify when we hear it. We understand the assumptions a Realist holds, as well as a liberalist, and it facilitates and mediates debates and discussions without the need for the need for extensive explanations of context, assumptions and terms. In other words, the "-isms" provide a structure that we are familiar with. Still, one wonders about the extent to which the "Great Debates" in IR have actually been resolved, or the pragmatism of even engaging them. Perhaps IR is better off assessing real-world problems and determining success under different criteria according to each approach (Lake, 2013).

Rathbun (2017), analyses the status of rationalism as a paradigm. This is not the classic classification, but it is both insightful and interesting for our purposes here. There seems a strong case for classifying the dominant paradigms of IR, especially Realism and Liberalism (including their myriad branches), as rationalist in their foundations. Whether or not this foundation itself should be classified as an IR tradition is not the issue here, it is enough that the debate is taking place. There should be little controversy in claiming that IR theory interprets human behaviour through the rationalist lens, with a "distinct logic of individualistic utilitarianism" (Rathbun, 2017).

The debate between rational choice and psychology, or rational choice versus behaviouralism is a tired one. While it is unclear when exactly an academic debate is clearly resolved, it seems (thanks largely to the behavioural scholars of the last century) that there is general agreement on the bounded nature of our rationality. Human beings, we agree on, are complicated, biased, emotional and rational creatures. All and none of these characteristics are true. It has been, and largely is, the task of scholars to specify under which circumstances we can expect these different responses.

There are significant limits to dominant IR paradigms. Indeed, the foundation upon which they construct their worldview is precarious. Arguably, they require deliberate and consistent updating. Earlier in this study I also referred to the tendency for the -isms to focus on system level variables in their analysis of events. Along with Lake (2013) and Rathbun (2017), I believe this part of their failing is because of the ideological taint in IR analysis. Tetlock (2005) notes the trend of political ideologues to fail most drastically in their political predictions. Strong ideological subscription, it seems, harms our ability to see problems in multivariate, complex and unbiased ways. The dominant IR theories are, arguably, ideologies as well as theories. They fit data – often, though not always – into theory, rather than adjusting theory for the data. We would do well to move beyond these theories, perhaps adapting them to insights from BIR and other fields, and adjusting our assessment of variables to include individuals and the decision-making process. The dominant paradigms have helped from our identity as IR scholars, now we might think of remaking that identity, or face the consequences of fading into redundancy.

So, for the reasons I have here outlined, traditional IR paradigms are not always particularly convincing in their analysis of bizarre IR phenomena. These phenomena precipitate a range of divisive narratives. Examples include the Donald Trump's rise to presidency, the Iraqi war of 2003 and Brexit. Narratives in both the academic literature and especially in the mainstream media, seem to present a simplistic view of the political world, sometimes backed by political commentators, who are often ideologically dominated. Given that such ideological possession harms prediction, we can be relatively sure that these explanations are inadequate. I have argued that these inadequacies present an opening for Behavioural IR to fill, and I have attempted to show that Behavioural IR – using the notion of cognitive bias as a core pillar – does indeed have the capacity to fill the gap.

2.3 In Pursuit of Simplicity

William of Ockham, a famous early 14th century philosopher stated that "pluralitas non est ponenda sine necessitate". That translated to "plurality should not be posited without necessity" (Duignan, 1999). It is an almost universally valid question to ask of any theoretical approach: Is this an overcomplication? Let us ask it of the Behavioural IR approach. It is a

misnomer, as far as I am concerned, to assume that the simplest explanation is the most correct one. Ockham certainly didn't think that was the case, and in any case, he is far from the last word on the topic. Some will no doubt claim that the ideological perspectives we seem to cling to in IR are quite simple explanations of behaviour – what with their assumptions of human nature clearly laid out and relatively simple. I think this harm, instead of helps, their case. It is important to view a phenomenon with the correct amount of complexity. Of course, things are often far more complicated than we can deal with – especially in the international political world. However, it would be wrong to try not to understand them better. I believe that BIR offers a multivariate microanalysis of events that promises much. It is possible to collect individual level data and apply it, perhaps with better predictive capacity than we currently possess (it can hardly be worse). The behavioural literature offers us an alternative to ideological possession that established a middle ground, a golden equilibrium. It has the potential to create parsimonious models that do not sacrifice complexity.

3. A Short Note on Interdisciplinarity

It is not our task here to delve into the issue of field divisions here too deeply. However, I want to iterate that the strict polychotomous nature of academia is an avoidable evil. The strict divisions between faculties, departments and projects severely restricts the creative multiplication of widely applicable ideas. The lack of integration between psychologists, economists and political scientists which led to the stunted propagation of behavioural work in International Relations is a small example of this shortcoming. Academics would do well to pursue integration between departments and faculties. Engineers need behavioural psychologists. Political scientists could learn a great deal from Physicists and Cosmologists. Indeed, some of the best investments made by the U.S. government have been related to the physical sciences. Comstock et al. (2011) note that efforts to quantify the benefits of the National Aeronautics and Space Administration's (NASA's) technology transfer efforts. Virtually all of the most significant and credible studies indicate that the U.S. economy has enjoyed a net economic and social benefit from government investments in NASA. For each dollar invested in NASA, the U.S. economy has reaped anywhere from seven to nine dollars

– a very favourable return (Comstock et al., 2011:3). Academics and policymakers could both be served by expanding their idea pool outside their own familiar fields.

4. Areas of Uncertainty

Most of the literature examined here above represent a collection of relatively well researched and understood biases and factors. There do remain however, some problematic areas, which need clarification for the research to move into the realm of functional knowledge.

How exactly do individual biases map onto group-level organisations or decision-making? This is not yet clear. Scholars have examined issues of groupthink at length, including Allison & Janis (1972, 1982), 't Hart (1990), Zelikow (1999), (Baron, 2005) and others. The literature on groupthink does not, however, clearly demonstrate its replicability in the laboratory, nor do we possess a reliable measure for the potential symptoms of the phenomena. We are still in the process of establishing testable hypothesis about the nature of group decision-making, which is perhaps the most useful contribution of the literature on groupthink. Another challenge facing behavioural analysts lies in determining the extent to which individuals and groups interact and influence outcomes. There does not yet exists a consistent measure of aggregating beliefs and decisions in a group, though efforts are being made in this direction (Russell, et al., 2015).

The study of individual political elites is important for behavioural analysis. In order to produce accurate data points, the collection of data on the personality profiles of contemporary leaders could lead to more accurate predictions of behaviour (O'Reilly, 2014; Wright & Tomlinson, 2018). With incentives for world leaders to engage in expiermental analysis low, it might be important to establish good proxies for gauging personality. While this approach will no doubt have its problems, it may yet provide important data for understanding the world of International Relations.

There exist a wide range of opportunities for scholars to engage in research that is both useful and interesting to them. The behavioural literature is – as a whole – rather young. As such there exists both a space and a need for the testing of previous hypotheses and the

identification of conditions for certain events. It is not enough to know that a particular bias exists. What will become increasingly useful is understanding the conditions for and specific nature of these biases, as well as corrective measures. Theories for corrective measures already exist, but also need testing, and this represents another avenue for scholars of BIR to explore.

5. Brief Summary

I will aim here to give a summarised account of my attempt to answer the main and sub questions which I presented at the beginning of the study. The main question asked if BIR explains bizarre events in the international world better than the dominant IR theories. Sub questions included: i) What is cognitive bias and how is it conceptualised in the literature? ii) What kind of events does traditional IR theory struggle to adequately explain? iii) How does BIR deal with these events? And iv) Does BIR provide a more compelling framework and analysis for understanding these so-called anomalous events?

The second chapter provided a theoretical picture of cognitive bias. It showed how cognitive bias stands at the centre of behavioural studies in IR. The concept of cognitive bias is significant in a wide range of fields, and studies highlighting its importance come from diverse fields. I attempted to outline a framework from which it was possible to understand the existence of different biases as well as their relationships with one another. It discussed the main insights from the literature on the evidence and implications of core biases, and it examined the different conceptualisations of the phenomena from the literature. I have discussed the issue at length in my analysis, and suffice it here to say that its role in IR should not be underestimated.

The third chapter analysed six texts in various related behavioural fields. It presented the key ideas and arguments of these texts. After presenting the notions surrounding cognitive bias in International Relations and illustrating the development of the field, a narrative synthesis was conducted. This analysis attempted to integrate the findings of the behavioural literature into a discussion which showed the evolution of ideas, terms and methods which constitute Behavioural International Relations. This discussion aimed to show the direction in which behavioural studies are heading in the future. By showcasing the state and promise of these

behavioural studies, a case is built for reassessing how we analyse International Relations phenomena by using notions of cognitive bias to inform models of behaviour.

Chapter four went on to examine some of the shortcomings of the core IR theories in terms of their ability to explain certain phenomena in the international world. I think it is a fair assumption that if ideologically dominated predictions fail, then accounts made in hindsight that resort to ideological explanans are likely to be incorrect. Instead, it might be possible that more fundamental variables are at play, namely, human psychology and bias. In chapter four I also outlined certain events which are not particularly well covered by traditional IR theory. I justified this claim by showing the predominant explanations from those points of view, which do not account for the events comprehensively. I showed how the literature identifies, characterises and uses cognitive bias and Behavioural IR theory to give a more comprehensive account for these events. Human psychology and bias, informational asymmetries and personality differences were all shown to play significant roles. Ultimately, I concluded that this type of analysis, which focuses more (though not exclusively) on individual-level variables provides more compelling explanations for behaviour.

Lastly, I show that my conceptualisation of cognitive bias is both useful and accurate. I give an account of the strength of my claims given the examined literature and conclude that it the account is a fairly robust one. Additionally, I posit that ideological simplicity harms the field of IR and that BIR presents a natural solution. Areas for future research are outlined, and are characterised by the need for clear scope conditions of cognitive biases, the need for analyses within experimentally valid environments and the need for a larger and more accurate collection of samples from political elites and current world leaders.

6. Accounting for (Mis)Behaviour and Bias

The inception of the preceding analysis began with a thought. Can we understand the human decision-making machine better? Language allows us to infer a surprising collection of information from certain events. By referring to phenomena with frames, the human-decision-maker (often unconsciously) categorises the relative importance of these phenomena. By referring to an event as bizarre, or shocking, we tend to frame the event as unexplorable. If we do not understand a phenomenon, the better approach is to assume that our current models of understanding are inadequate – and therefore, that we need to update them. It is the easier, and the less productive, path to dismiss events as opposed to analysing them and attempting to make sense of them. This review discussed the possibility that outcomes labelled as bizarre may in fact be fairly congruent with a different kind of political analysis – that is, the type of analysis that BIR can offer. This review attempted to demonstrate the opportunity for more accurately defining and explaining baseline behaviour and enhancing our understanding of the causes, nature and results of cognitive bias in IR.

Human beings have been shown to fall short of rational-model assumptions. They make predictable errors in judgement based on cognitive biases. However, the most popular theoretical lenses in International Relations have not adapted their models for this reality, and thus misunderstand behaviour that is – promisingly – understandable, albeit with a different theoretical lens. By using BIR as this lens, it might be possible to modify our collective understanding of events in International Relations, which would contribute to establishing and maintaining peace, stability and cooperation in the international order. I have attempted to construct an answer to the question: In light of "anomalous" behaviour by actors in IR, often reduced to irrationality by traditional IR theory, might BIR – which focuses on cognitive bias as a core variable in its analysis – better explain such behaviour?

I hypothesised that this integrative review will find that cognitive bias and cognition are key explanatory variables in grasping anomalous behaviour by actors in International Relations. I think that I have made a case that is strong enough to confirm this hypothesis.

I started building my case by assessing the state and nature of cognition and cognitive bias in International Relations. The lack of a unified theory of cognition and cognitive bias in IR prompted me to build a simple model which captured the most important processes and

biases which I identified in the literature. I outlined some of the key theoretical literature, which argues that under certain conditions, decision-makers are influenced more strongly by emotive, psychological and biological factors than they are by strategic gain. Individual preferences and beliefs matter greatly to the human brain, often more than questions of net gains. Under this behavioural approach, a 2-stage model of cognition has been widely accepted. This model distinguishes between intuitive, instantaneous processes and rational, deliberative ones. The patterns of biases and heuristics that characterise human thinking vary predictably depending on which mode of cognition one uses in a given scenario.

Other theories break down errors in thinking into categories depending on the biological foundations of the errors. Other theories categorise these errors by viewing them as the culmination of cognitive processes. This theoretical foundation then justified the conceptual framework I created to describe the human cognitive process. This framework described a process that included: i) reception of information, which included the manner in which information is accumulated and stored ii) perception of information, which referred to the way information was analysed and processed and iii) decision, which denoted the act of weighing and measuring the processed evidence to conclude about which the correct course of action is. The Decision step in my conceptualisation was also shown to be heavily influenced by the decision-makers preferences and beliefs.

The proposed framework for cognitive bias represents an important phenomenological claim. We do not, intuitively, have complete access to the external world, upon experiencing it we immediately store information in an imperfect way and change it. However, the manner in which we interpret reality are not – in practice – infinite. Rather, they are predictable across people. We can use this insight to i) better predict and understand human behaviour and ii) reconcile varying views based on their (almost inevitable) commonalities.

After giving an account of cognition and positioning it within this framework, I distilled the most important insights form important texts for Behavioural International Relations. I used an integrative review method, beginning by distilling information from six texts into a comprehensive and comprehensible summary, grounding the reader in the most important contributions of the literature to the ideas of cognitive bias and cognition. Secondly, I formed

a discussion in order to tease out the evolution of concepts, methods and approaches of behavioural scientists in IR.

Behaviouralism in Political Science has, in some cases been contrasted with rational choice. In other cases, it has played a complementary role to rational choice models, by updating models of actor preference, beliefs and the cognitive process. In the area of emotions, more fundamental differences are at play between these two schools of thoughts. Since the first behavioural revolution, more than half a century ago, significant contributions to theory and research have been made. In moving forwards with this research, scholars have begun to identify the scope conditions for both standard and nonstandard assumptions about actor behaviour. The experimental nature of contemporary scholarship gives some scholars hope that this new behavioural revolution will disperse more widely into IR. Other scholars in IR are wary of the rifts that could open up between "deductively valid theories and empirical research in international relations" (Hafner-Burton et al., 2017).

My analysis finished by noting that behavioural scholarship in International Relations over the last fifty years has been insightful and useful. Future research will need to synthesise a more comprehensive and unified theory of behaviour in IR. If it does, it may constitute some of the most relevant work political scientists have conducted in recent years.

After showing how the literature deals with the concepts at hand, and asserting the strengths of their approaches, I pushed my analysis further. I dealt with (mis)behaviour and bias by demonstrating that there are certain phenomena that are poorly explained by traditional IR theory, and that these phenomena can be more satisfactorily explained by the methods and concepts that I identified earlier on in the review. These phenomena included (non)compliance in IR, beliefs in policymaking, the invasion of Iraq, the forecasting inaccuracy of political experts and climate change negotiations.

On the weight of the evidence, it seems reasonable to conclude that the 'misbehaviour' of actors in International Relations is, in fact, merely behaviour. We already possess sophisticated enough tools to make sense of these anomalous events, these tools are part of the behavioural toolkit. Traditional IR theories need, at a minimum, to respond to the evidence that this review uncovers by updating their respective worldviews. In some cases, updating will be sufficient to significantly improve their explanatory power and more

correctly perceive the reality of international politics. This integrative review succeeds in renewing the call for scholars to consider the value which the methodological and theoretical foundations of Behaviouralism and BIR add to International Relations. Pure rational choice models have fundamental limitations. Judgement only partly explains our behaviour. Attitudes and beliefs are important predictors and determinants of behaviour. BIR needs to establish a more unified theoretical framework, conduct experiments in high-stakes environments with political elites and deal with the issue of how individuals and collective entities mitigate, exacerbate or otherwise affect cognitive biases.

Bibliography

Allison, G. T. & Zelikow, P., 1999. *Essence of Decision: Explaining the Cuban Missile Crisis.* 2 ed. New York: Longman.

Antunes, S. & Camisao, I., 2017. Realism . In: S. McGlinchey, R. Walters & C. Scheinpflug, eds. *International Relations Theory.* Bristol, UK: University of West England, pp. 15-21.

Ardanaz, M., Murillo, M. V. & Pinto, P. M., 2013. Sensitivity to Issue Framing on Trade Policy Preferences: Evidence from a Survey Experiment. *International Organisation ,* 67(2), pp. 411-437.

Århem, P. & Liljenströmb, H., 1997. On the Coevolution of Cognition and Consciousness. *Journal of Theoretical Biology,* 187(4), pp. 601-612.

Arkes, H. R., 1991. Costs and Benefits of Judgement Errors: Implications for Debiasing. *Psychological Bulletin,* 110(3), pp. 486-498.

Baron, I. Z., 2014. The Continuing Failure of International Relations and the Challenges of Disciplenary Boundaries. *MILLENNIUM Journal of International Studies,* 43(1), pp. 224-244.

Baron, R. S., 2005. So right it's wrong: Groupthink and the ubiquitous nature of polarized group. *Advances in experimental social psychology,* 37(2), pp. 219-253.

Bénabou, R., 2015. The economics of motivated beliefs. *Revue d'économie politique,* 125(5), pp. 665-685.

Berkenpas, J. R., 2016. *The Behavioral Revolution in Contemporary Political Science: Narrative, Identity, Practice,* Michigan: Western Michigan University.

Boell, S. K. & Cecez-Kecmanovic, D., 2014. A Hermeneutic Approach for Conducting Literature Reviews and Literature Searches. *Communications of the Association for Information Systems,* 34(12), pp. 257-286.

Bryman, A., 2012. *Social Research Methods.* 4th ed. New York: Oxford University Press.

Camerer, C. F., Ho, T.-H. & Chong, J.-K., 2004. A Cognitive Hierarchy Model Of Games. *The Quarterly Journal of Economics,* August(2004), pp. 861-898.

Chai, S.-K., 2001. The success and failure of rational choice. In: A. Arbor, ed. *Choosing an Identity: A General Model of Preference and Belief Formation.* s.l.:University of Michigan Press, pp. 1-20.

Checkel, J. T., 1998. Review: The Constructivist Turn in International Relations Theory. *World Politics,* 50(2), pp. 324-348.

Colaresi, M., 2004. When doves cry: International rivalry, unreciprocated cooperation, and leadership turnover. American Journal of Political Science, 48(3), pp.555-570.

Comstock, D.A., Lockney, D. and Glass, C., 2011. A Sustainable Method for Quantifying the Benefits of NASA Technology Transfer. In AIAA SPACE 2011 Conference & Exposition (p. 7329).

Dahl, R., 1961. The Behavioral Approach in Political Science: Epitaph for a Monument to a Successful Protest. *The American Political Science Review,* 55(4), pp. 763-772.

David, C.-P. & Prémont, K., 2015. *Bad Analogical Reasoning and Post-War Operations in Iraq After 2003.* San-Diego, CA: Western Political Science Association (WPSA).

Dawisha, A. et al., 2013. Review: Ideology, Realpolitik, and US Foreign Policy: A Discussion of Frank P. Harvey's. *Perspectives on Politics,* 11(2), pp. 578-592.

DellaVigna, S., 2009. Psychology and Economics: Evidence from the Field. *Journal of Economic Literature,* 47(2), pp. 315-372.

Dietrich, F. and List, C., 2016. Mentalism versus behaviourism in economics: a philosophy-of-science perspective. Economics & Philosophy, 32(2), pp.249-281.

Duignan, B., 1999. *Occam's razor.* [Online] Available at: https://www.britannica.com/topic/Occams-razor#accordion-article-history. [Accessed 04 August 2018].

Duune, T., Hansen, L. & Wight, C., 2013. The end of International. *European Journal of International Relations,* 19(3), pp. 405-425.

Dyson, S. B., 2006. Personality and Foreign Policy: Tony Blair's Iraq Decisions. *Foreign Policy Analysis,* 2(1), pp. 289-306.

Easton, D., 1962. The current meaning of “Behavioralism” in political science. In: 1, ed. *The limits of behavioralism.* Philadelphia: American Academy of Political and social science1, pp. 8-25.

Eldridge, A., 1977. Reviewed Work: Perception and Misperception in International Politics by Robert Jervis. *The Journal of Politics,* 39(4), pp. 1106-1108.

Elman, C., 2007. Realism. In: M. Griffiths, ed. *International Relations Theory for the Twenty-First Century: An introduction.* New York, NY: Routledge, pp. 11-20.

Farnham, B., 1994. *Avoiding Losses/Taking Risks: Prospect Theory and International Conflict.* 1st ed. Michigan: University of Michigan Press.

Ferguson, Y.H., 2015. Diversity in IR theory: Pluralism as an opportunity for understanding global politics. International Studies Perspectives, 16(1), pp.3-12.

Ferreira, M. F., 2017. Critical Theory . In: S. McGlinchey, ed. *International Relations Theory.* Bristol, UK: University of the West of England , pp. 49-55.

Foreman, E. & Selly, M., 2001. Decision by Objectives. Singapore: World Scientific.

Fujii, G. ISSF Roundtable 10-4 on Perception and Misperception in International Politics and on How Statesmen Think: The Psychology of International Politics. H-Diplo. 12-08-2017. https://networks.h-net.org/node/28443/discussions/1000429/iss-roundtable-10-4-perception-and-misperception-international

Frederiks, E., Stenner, K. & Hobman, E., 2015. Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour. *Renewable and Sustainable Energy Reviews,* Volume 41, pp. 1385-1394.

Golden-Biddle, K. & Locke, K. D., 2007. *Composing Qualitative Research.* 2nd ed. London: Sage Publications.

Golman, R., Hagmann, D. and Loewenstein, G., 2017. Information avoidance. Journal of Economic Literature, 55(1), pp.96-135.

Griffiths, M., 2007. *International Relations Theory for the 21st Century: An Introduction.* 1 ed. New York: Routledge.

Haas, M., 2016. *International Relations Theory: Competing Empirical Paradigms.* 1 ed. New York: Lexington Books.

Hafner-Burton, E. M., Haggard, S., Lake, D. A. & Victor, D. G., 2017. The Behavioral Revolution and International Relations. *International Organisation,* 71(1), pp. 1-31.

Hafner-Burton, E. M., Hughes, D. A. & Victor, D. G., 2013. The Cognitive Revolution and the Political Psychology of Elite Decision-making. *Perspectives on Politics,* 11(2), pp. 368-386.

Hafner-Burton, E. M., LeVeck, B. L. & Victor, D. G., 2017. False Promises: How the Prospect of Non-Compliance Affects. *International Studies Quarterly,* 61(1), pp. 136-149.

Hafner-Burton, E. M., LeVeck, B. L., Victor, D. G. & Fowler, J. H., 2014. Decision Maker Preferences for International Legal Cooperation. *International Organization,* 68(Fall), pp. 845-876.

Hafner-Burton, E. M., Leveck, B. & Victor, D. G., 2015. *How The Prospect of Non-Compliance Affects Elite Preferences for International Cooperation: Evidence From a "Lab in the Field" Experiment.* San Diego, CA: Laboratory on International Law and Regulation (ILAR).

Harvey, F. P., 2011. *Explaining the Iraq War: Counterfactual Theory, Logic and Evidence.* 1 ed. New York: Cambridge Universtity Press.

Hellmann, G., 2011. International Relations as a field of study. In: 1, ed. *International Encyclopedia of Political Science.* Thousand Oaks, CA: SAGE Publications, pp. 1298-1315.

Hovi, J., Detlef, S. F. & Bang, G., 2012. Why the United States Did Not Become a Party to the Kyoto Protocol: German, Norwegian and US Perspectives. *European Journal of International Relations,* 18(1), pp. 129-150.

Ilker Etikan, Sulaiman Abubakar Musa, Rukayya Sunusi Alkassim. Comparison of Convenience Sampling and Purposive Sampling. American Journal of Theoretical and Applied Statistics. Vol. 5, No. 1, 2016, pp. 1-4

Jack, L. 1997. Prospect Theory and the Cognitive-Rational Debate. In Decisionmaking on War and Peace: The Cognitive-Rational Debate, edited by Nehemia Geva and Alex Mintz. Boulder: Lynne Rienner Publishers. James, P., 2007. Behavioral IR: Practical Suggestions. *International Studies Review,* 9(1), pp. 162-165.

Jack, L. 2003. Political Psychology and Foreign Policy. In Oxford Handbook of Political Psychology, edited by David O. Sears, Leonie Huddy, and Robert Jervis. New York: Oxford University Press.

Janis, I. L., 1972. *Victims of groupthink: a psychological study of foreign-policy decisions and fiascoes..* 1st ed. Oxford, England: Houghton Mifflin.

Janis, I. L., 1982. *Groupthink: Psychological Studies of Policy Decisions and.* 1st ed. Boston: Houghton Mifflin.

Jervis, R., 1976. *Perception and Misperception in International Politics.* 1st ed. Princeton, New Jersey: Princeton University Press.

Jervis, R., 2017. Perception and Misperception in International Politics: New Edition. Princeton University Press.

Kahneman, D., 2003. Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review,* 93(5), pp. 1449-1475.

Kahneman, D., 2012. *Thinking, Fast and Slow.* 2nd ed. London: Penguin.

Kahneman, D., Knetsch, J. L. & Thaler, R. H., 1991. Anomalies. *Journal of Economic Perspectives ,* 5(1), pp. 193-206.

Kahneman, D. & Tversky, A., 1979. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica,* 47(2), pp. 263-292.

Kahneman, D. & Tversky, A., 1984. Choices, Values, and Frames. *American Psychologist,* 4(39), pp. 341-350.

Karp, L., 2004. Global warming and hyperbolic discounting. *Journal of Public Economics,* 89(2005), pp. 261-282.

Katz, J., 2015. A Theory of Qualitative Methodology: The Social System of Analytic Fieldwork. *Méthod(e)s,* 1(1 & 2), pp. 131-146.

Kertzer, J. D. & Tingley, D., 2018. Political Psychology in International Relations: Beyond the Paradigms. *Annual Review of Political Science,* 21(1), pp. 1-23.

Lake, D. A., 2011. Two Cheers for Bargaining Theory: Assessing Rationalist Explanations of the Iraq War. *International Security,* 35(3), pp. 7-52.

Lake, D. A., 2013. Theory is dead, long live theory: The end of the Great Debates and the rise of eclectism in International Relations. *European Journal of International Relations ,* 19(3), pp. 567-587.

Lichtenstein, S. & Slovic, P., 2006. *The Construction of Preference.* 1 ed. Cambridge, UK: Cambridge University Press.

Mearsheimer, J. J. & Walt, S. M., 2003. Leaving theory behind: Why simplistic hypothesis testing is bad for International Relations. *European Journal of International Relations ,* 19(3), pp. 427-457.

Meiser, J. W., 2017. Liberalism . In: S. McGlinchey, ed. *International Relations Theory .* Bristol, UK: University of West England , pp. 22-27.

Mellers, B. et al., 2015. The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics. *Journal of Experimental Psychology:Applied,* 21(1), pp. 1-14.

Miller, J.L., Cramer, J., Volgy, T.J., Bezerra, P., Hauser, M. and Sciabarra, C., 2015. Norms, behavioral compliance, and status attribution in international politics. International Interactions, 41(5), pp.779-804.

Mintz, A., 2004a. Foreign policy decision making in familiar and unfamiliar settings: An experimental study of high-ranking military officers. Journal of Conflict Resolution, 48(1), pp.91-104.

Mintz, A., 2004b. How Do Leaders Make Decisions? : A Poliheuristic Perspective. *Journal of Conflict Resolution,* 48(3), pp. 1-13.

Mintz, A., 2007. Behavioral IR as a Subfield of International Relations. *International Studies Review,* 9(1), pp. 157-162.

Mintz, A. & DeRouen Jr, K., 2010. *Understanding Foreign Policy Decision-making.* 1 ed. Cambridge, UK: Cambridge University Press.

Mintz, A. & Redd, S. B., 2003. Framing Effects in International Relations. *Synthese,* 135(1), pp. 193-213.

Montano, D. E. & Kasprzyk, D., 2015. Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. In: K. Glanz, B. K. Rimer & K. Viswanath, eds. *Health Behavior: Theory, Research, and Practice.* Hoboken, New Jersey: John Wiley & Sons, pp. 95-124.

Montibeller, G. & von Winterfeldt, D., 2015. Cognitive and Motivational Biases in Decision and Risk Analysis. *Risk Analysis ,* 35(7), pp. 1230-1251.

Mousavi, S., Gigerenzer, G. & Kheirandisch, R., 2016. Rethinking Behavioral Economics through Fast-and-Frugal Heuristics. In: F. R, et al. eds. *Routledge handbook of behavioral economics.* London : Taylor & Francis, pp. 280-296.

Nhandara, S., 2015. *The Operational Code of Tony Blair: Did he experience Learning, Stability or Change in his Belief Systems during the period he was Prime Minister?.* Stockholm: Södertörn University.

Nincic, M., 1997. Loss aversion and the domestic context of military intervention. Political Research Quarterly, 50(1), pp.97-120.

O'Reilly, K. P., 2014. *Nuclear Proliferation and the Psychology of Political Leadership: Beliefs, Motivations and Perceptions.* 1 ed. New York and London: Routledge.

Overall, J., 2016. The Dark Sise of Entrepreneurship: A Conceptual Framework of Cognitive Bias, Neutralization, and Risky Entrepreneurial Behavior. *Academy of Entrepreneurship Journal ,* 22(2), pp. 1-12.

Pentony, D., 2005. Supplementary Reading Summaries, San Francisco: San Francisco State University.

Peterson, G., 2017. *The Effects of System Justification and Social Dominance Orientation on the Temporal Discounting of Climate Chang.* Orange, CA: Chapman University.

Post, J. M., 1991. Saddam Hussein of Iraq: A Political Psychology Profile. *Political Psychology,* 12(2), pp. 279-289.

Puchala, D. J., 2003. *Theory and History in International Relations.* 1 ed. New York and London: Routledge.

Rabin, M., 1998. Psychology and Economics. *Journal of Economic Literature ,* 36(1), pp. 11-46.

Rathbun, B. C., 2017. Subvert the dominant paradigm: A critical analysis of rationalism's status as a paradigm of international relations. *International Relations,* 31(4), pp. 403-425.

Regenwetter, M., Dana, J. and Davis-Stober, C.P., 2011. Transitivity of preferences. Psychological Review, 118(1), p.42.

Robison, B. S., 2011. "Experiential Learning by US Presidents: Domestic, International Influences in the Post-Cold War World. In: Walker, G, S. Malici, A. and Schafer, M., ed. *Rethinking Foreign Policy Analysis: States, Leaders, and the Micro foundations of Behavioral International Relations.* New York: Routledge.

Rosenau, James N, 1965. Behavioral science, behavioural scientists and the study of international phenomena: a review. The Journal of Conflict Resolution, 9(4), pp. 509-521.

Russell, J. S., Hawthorne, J. & Buchak, L., 2015. Groupthink. *Philosophical studies,* 172(5), pp. 1287-1309.

Schwarz, N., 2000. Emotion, cognition, and decision-making. *COGNITION AND EMOTION,* 14(4), pp. 433-440.

Sheppard, B. H., Hartwick, J. & Warshaw, P. R., 1988. The Theory of Reasoned Action: A Meta-Analysis of Past Research with Recommendations for Modifications and Future Research. *Journal of Consumer Research,* 15(3), pp. 325-343.

Simmons, B. A., 1998. Compliance with international agreements. *Annual Review of Political Science,* 1(1), pp. 75-93.

Simon, H., 1985. Human Nature in Politics: The Dialogue of Psychology with Political Science. *The American Political Science Review,* 79(2), pp. 293-304.

Slaughter, A.-M., 2011. International Relations, Principal Theories. In: R. Wolfrum, ed. *Max Planck Encyclopedia of Public International Law.* Oxford, UK: Oxford University Press.

Smith, S., Dunne, T. & Hadfield, A., 2016. *Foreign Policy: Theories, Actors, Cases.* 3rd ed. Oxford, UK: Oxford University Press.

Stein, J. G., 2017. The Micro-Foundations of International Relations Theory: Psychology and Behavioral Economics. *International Organization ,* 71(1), pp. 249-263.

Sternberg, R. J., Sternberg, K. & Mio, J., 2012. *Cognitie Psychology.* 6th ed. Belmont, CA: Wadsworth.

Stivachtis, Y. A., 2017. The English School. In: S. McGlinchey, ed. *International Relations Theory.* Bristol, UK: University of the West of England, pp. 28-35.

't Hart, P., 1990. *Groupthink in government: A study of small groups and policy failure.* 1st ed. Lisse, Netherlands: Swets & Zeitlinger Publishers.

Tetlock, P. E., 2005. *Expert Political Judgement.* 1 ed. New Jersey: Princeton University Press.

Thaler, R. H. & Ganser, L. J., 2015. *Misbehaving: The making of behavioral economics.* 1 ed. New York, NY: WW Norton.

The Guardian, 2003. Full text of Colin Powell's speech. [Online]. Available at: https://www.theguardian.com/world/2003/feb/05/iraq.usa. [Accessed 20 10 2018].

Theys, S., 2017. Constructivism . In: S. McGlinchey, ed. *International Relations Theory.* Bristol, UK: University of the West of England , pp. 36-41.

Trautmann, S. T. & Zeckhauser, R. J., 2010. *Blindness to the benefits of ambiguity: the neglect of learning opportunities..* Cologne, IAREP/SABE/ICABEEP 2010 Conference.

Tschoegl, A. E. & Armstrong, S. J., 2007. Review of: Philip E. Tetlock. 2005. Expert Political Judgment: How Good Is It? How Can We Know?. *International Journal of Forecasting,* 23(2), pp. 339-342.

Tversky, A. & Kahneman, D., 1974. Judgment under Uncertainty: Heuristics and Biases. *Science,* 185(4157), pp. 1124-1131.

United Nations, UNOHCHR, 2016. *Ratification of 18 International Human Rights Treaties.* [Online]. Available at: http://indicators.ohchr.org. [Accessed 24 October 2018].

Victor, L., 2008. Systematic Reviewing in the Social Sciences: Outcomes and Explanation. *Social Research Update,* 1(54), pp. 1-4.

Von Eckardt, B., 1995. *What is Cognitive Science.* 1st ed. Cambridge, Massachusetts: The MIT Press.

Waever, O., 2013. Still a Discipline after All These Debates?. In: T. Dunne, M. Kurki & S. Smith, eds. *International Relations Theories: Discipline and Diversity.* Oxford: Oxford University Press, pp. 306-328.

Walker, S. G., 2007. Back to the Future? Behavioral IR as a Case of Arrested Development. *International Studies Review,* 9(1), pp. 165-170.

Walt, S. M., 1998. International relations: One world, many theories. *Foreign Policy,* Issue 110, pp. 29-35.

Waltz, K., 1959. *Man, the State and War: a Theoretical Analysis.* New York and London: Columbia University Press.

Waltz, K. N., 1979. *Theory of International Politics.* 1 ed. Michigan: McGraw-Hill.

Whittemore, R. and Knafl, K., 2005. The integrative review: updated methodology. Journal of advanced nursing, 52(5), pp.546-553.

Wogu, I. A. P., 2013. Behaviouralism As An Approach To Contemporary Political Analysis: An Appraisal. *International Journal of Education and Research,* 1(12), pp. 1-12.

Wright, J. D. & Tomlinson, M. F., 2018. Personality profiles of Hillary Clinton and Donald Trump: Fooled by your own politics. *Personality and Individual Differences,* 128(1), pp. 21-24.

Zhang, H.-B., Dai, H.-C., Lai, H.-X. & Wang, W.-T., 2017. U.S. withdrawal from the Paris Agreement: Reasons, impacts, and China's response. *Advances in Climate Change Research,* 4(8), pp. 220-225.

Zinnes, D. A. (1978) “Book Review: Perception and Misperception in International Politics,” The American Political Science Review, 72(2), pp. 793–794.