

USING PSYCHOLOGICAL INDICATORS TO PREDICT CRISIS ESCALATION: A
TRACKING APPROACH

by

MICHAEL ROWLAND STEWART

M.A., The University of St Andrews, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

The Faculty of Graduate Studies

(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

AUGUST 2009

© Michael Rowland Stewart, 2009

ABSTRACT

This project investigated the role that tracking psychological indicators (measured using the content analysis of political rhetoric) could play in predicting political behaviour during a political crisis. An initial review of the literature revealed that a number of psychological variables seem to be related to violence and/or co-operation. These included power motivation, complexity, belief in ability to control events, and belief about the cooperativeness of the political world. Hence, all of these variables were tested for their utility in predicting political aggression/cooperation during the crisis in Zimbabwe in an historical analysis that extended from the beginning of 2007 to the end of 2008. A time-series regression analysis revealed that a large proportion of government violence during the 2007/2008 period could be accounted for by the complexity (measured through integrative complexity), power motivation (measured through power motive imagery) and belief in ability to control events of Nathaniel Manheru, a columnist for a state newspaper and close aide of Zimbabwe President Robert Mugabe. Integrative complexity and power motive imagery in particular, seemed to provide unique and powerful predictive utility in this model. However, given that no out-of-sample forecasting was possible in this application study, uncertainty remains with regard to the use of such a model for real-time forecasting – a problem that can be rectified in future research.

TABLE OF CONTENTS

Abstract	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
Acknowledgements	vii
Dedication	viii
1 Introduction	1
1.1 The Problem of Predicting Crisis Behaviour	1
1.2 Psychological Factors Associated with Crisis Behaviour: A Review	3
1.2.1 Power motivation/need for power	3
1.2.2 Complexity	6
1.2.3 Optimism/belief in ability to control events	11
1.2.4 Beliefs about the political world/distrust	13
1.2.5 Summary of psychological indicators of intent	15
1.3 Applying Psychological Factors to Predict Aggression/Cooperation	16
1.3.1 Profiling approach	16
1.3.2 Tracking approach	24
1.4 A Study Testing the Utility of the Tracking Approach to Prediction.....	25
2 Method	29
2.1 Selection of a Crisis Test Case and Scoring Materials.....	29
2.2 Independent Variables	31
2.2.1 Complexity - Integrative complexity.....	32
2.2.2 Complexity - Conceptual complexity	33
2.2.3 Power motivation – Power motive imagery	35
2.2.4 Power motivation - Need for power	36
2.2.5 Belief in ability to control events.....	37
2.2.6 Philosophical beliefs about the political others (P1).....	38
2.3 Dependent Variables.....	39
2.3.1 Conflict and co-operation by the Zimbabwe Government.....	39
2.4.1 Conflict and co-operation events	42
2.4.2 Psychological indicators – Computer scored.....	44
2.4.3 Psychological variables – Hand scored	44
2.4.4 Time aggregation.....	45
3 Results	46
3.1 Initial Data Analysis	46

3.2 Fitting a Model to the Data	49
3.3 Out of Sample Forecasting.....	51
4 Discussion.....	58
4.1 Evaluation of the Central Findings	58
4.1.1 Question 1.....	58
4.1.2 Question 2.....	60
4.1.3 Questions 3 and 4	62
4.1.4 Question 5.....	63
4.2 Further Issues, Problems and Criticisms	69
4.2.1 Weaker than normal reliability	69
4.2.2 Practical considerations	70
4.2.3 Other approaches to predicting conflict: Where a tracking approach fits in.....	71
4.3 Conclusion - The Tracking Approach to Predicting Adversarial Intent.....	73
References	74

LIST OF TABLES

Table 3.1	52
Table 3.2	53
Table 3.3	54
Table 3.4	55

LIST OF FIGURES

Figure 3.1	55
Figure 3.2	56
Figure 3.3	56
Figure 3.4	57

ACKNOWLEDGEMENTS

I offer my deepest gratitude to all those who have helped me through the last two years.

In particular, I would like to thank the following individuals, without whom this research would not have been possible:

Dr Peter Suedfeld

Ryan Cross and Sverre Frisch

Mariel Lopa, Dora Tsao, Manpreet Narwal

Katya Legkaia, Jelena Brcic

Lastly, I would also like to thank Defence Research and Development Canada who funded this research under Contract W7711-07-8104 (P. Suedfeld, Principal Investigator).

To Jen and my parents

1 INTRODUCTION

1.1 The Problem of Predicting Crisis Behaviour

History has shown that during political crises, when the fate of entire countries may be at stake, political judgment and decision-making are crucial. One only needs to look at the Cuban Missile Crisis in 1962 or the Munich Crisis prior to the Second World War, for instance, to see that it is during times such as these that a small number of political leaders can hold the future of an entire country or region in their hands. It has equally been seen during times such as these that political behaviour has been inherently unpredictable. Many thought the world would end during the Cuban Missile Crisis in 1962, whilst Neville Chamberlain believed he had secured “peace for” his “time” at the Munich agreement in 1938.

Thus, because of the of the unpredictable nature of political crises and the high stakes that are inevitably involved, being able to predict political behaviour more accurately during these times should be of enormous interest to us all. Doing so can not only enhance our ability to detect genuine adversarial intent in our opponents but also enhance our ability to disregard brinkmanship and bluff. Perhaps many crises could be avoided all together. If psychology, or any field for that matter, could be effectively applied to predict political behaviour in these circumstances (or even provide accurate current assessments of those circumstances) this would be a most useful political application.

It is helpful therefore that political psychology may have much to offer in this regard. For the past 40 years, political psychologists have been studying the personalities (Hermann, Preston, Korany & Shaw, 2001; Hermann, 2006), motives (Winter, 2005; 2006; Winter & Stewart, 1977), cognitive style (Suedfeld, Guttieri & Tetlock, 2006) and beliefs (Walker, Schafer & Young,

1998; 2006) of leaders by quantitatively assessing the properties of those leaders' spoken language. In doing so, they have found that measurement of these psychological dimensions can provide solid and consistent explanations of those leaders' political behaviour (see Winter, 2003b, for summary). Crucially, included in this repertoire of explainable behaviours are acts of aggression (e.g., ordering an attack on one's political opponents) and acts of co-operation (e.g., agreeing to meet or negotiate), both of which are of primary interest when predicting behaviour during a political crisis (Suedfeld, Cross & Stewart; 2008).

Given the importance of predicting decision-making during political crises, the question naturally arises of how one might approach the task of predicting behaviour during political crises from a psychological point of view. Is it possible to use psychological measurements to predict aggression in political opponents in a systematic and reliable manner? Could these methods provide the kind of indicators of aggressive or co-operative intent that are so badly needed during periods of political instability?

These are compelling questions, so it is unfortunate then that little political psychological research to date has attempted to speak to the question of predictive applications directly. Even the small amount of research that *has* been explicitly concerned with predicting specific leaders' behaviour (e.g., Winter et al., 1991a, Winter, 2001; see below) has not been focussed specifically on predicting behaviour during political crises but more general behavioural tendencies in a diverse set of circumstances.

The purpose of this thesis therefore is to begin filling this gap in the literature. It does this by first exploring *theoretically* how the robust research findings in political psychology – those findings that link psychological factors to political aggression or co-operation - might be best applied to predict such behaviour during political crises. This includes a discussion on how real predictions

using an established approach (i.e., the profiling approach to prediction) have performed so far. It then tests a new approach (i.e., a tracking approach to prediction) *empirically* using a detailed case study – the political crisis in Zimbabwe.

However, before reviewing the different approaches one might use to harness psychological factors for the purpose of prediction during political crises, it is first appropriate to review the psychological factors that could be fitting for that purpose. What psychological factors are related to aggression and co-operation and, moreover, which relationships between psychological indicators and aggression have been found to be robust? In terms of answering the question of the predictive/associative *relationship* between psychological indicators and political aggression/co-operation (i.e., as opposed to the predictive *application* of psychological indicators) political psychology research has certainly not fallen short.

1.2 Psychological Factors Associated with Crisis Behaviour: A Review

There is now a wealth of research into the personality, motives, cognition and beliefs of leaders that link psychological indicators with aggression and co-operation in political crises. Such indicators could, if applied prudently, also be the tools by which political psychologists go on to make sound predictions about crisis behaviour. Each is now described here in turn, along with the evidence supporting their use in a psychological model of adversarial intent.

1.2.1 Power motivation/need for power

A string of studies by David Winter (2005; 2003b; 1973), and his colleagues on motives (see Smith, Atkinson, McClelland, & Veroff, 1992) has shown that typically those leaders who are concerned with obtaining or maintaining power or prestige, or having control or impact on others, are also more likely to use aggressive political methods. For instance, Winter (1987b) has reported that across a number of U.S. presidents the presence of power motive imagery in the

president's language is moderately and positively correlated with war entry during the President's term, whilst a higher level of power imagery in annual British Sovereigns' speeches has been found to precede British entry into war in the following year (Winter, 1993).

This pattern of findings also extends beyond the Western world. For instance, Winter's (1980) examination of the motives in southern African leaders has revealed that power motivation is strongly related to independent judges' ratings of African leaders' war dispositions (i.e., their likelihood of entry into war) whilst across a number of international crises with different actors, Winter (2007) has shown that imagery associated with power motivation is comparatively ubiquitous in inter-government communications, speeches and broadcasts before crises ending in conflict (e.g., the Bay of Pigs Disaster) versus those ending in peaceful resolution (e.g., the Cuban Missile Crisis), regardless of the states involved.

Lastly, two studies by Winter (2003a; 1987a) have found that in addition to the general presence of power imagery in communications before conflict, there is also an asymmetry in the power imagery depending upon which side is being referenced. Communications preceding descent into severe crises evidenced an enhanced degree of power imagery when they were in reference to the opposition but not when they were in reference to the communicator's home government. Interestingly, this asymmetry was not constant across time but occurred only when the crises reached tipping point. Moreover, it was found not just in the media and public communications but in private communications as well - even in the communications expressed by those individuals who unambiguously desired a peaceful outcome.

Other research has replicated Winter's general findings but in different settings or with different methodology – an indicator of just how robust this relationship is. For example, Smith (2008) has discovered that power motivation is related to a preference for violence in groups working at

the sub-state level. She found that the use of power imagery is significantly greater amongst terrorist groups than non-terrorist groups even when these groups have similar ideologies. Meanwhile, Keller (2005a; 2005b) has replicated Winter's findings with an entirely different method scoring for power motivation - Hermann's 'need for power' (see Hermann, 2006, 1980, Winter, 1973, Smith, Atkinson, McClelland, & Veroff. 1992). Keller found that the relative *constraint challenging* natures (of which 'need for power' is a key component) of thirty-nine different world leaders were strongly related to the likelihood of them using violence during a political crisis. They discovered that there was nearly a fifty percent chance that leaders high on the constraint-challenging dimension had used violence as the pre-eminent response to an international crisis, whilst those who scored lowest on this dimension almost never used violence at the first moment of opportunity. This built on similar earlier work by Hermann (1980), who found a moderate correlation between need for power and her measure of policy aggressiveness.

Lastly, Dyson (2006; 2007), who also measured power motivation using Hermann's method, has proposed that a concern for power may have been a key factor behind the former British Prime Minister (PM) Tony Blair's decision to join the U.S. in the 2003 Iraq War. Dyson (2006) found that Blair was much higher than other British Prime Ministers on the 'need for power' dimension, indicating that the PM was exceptional in this respect. In contrast, he also discovered that the former British PM Harold Wilson, who faced a very similar decision with regard to entering the Vietnam War (1965-1975), had been decidedly average in his need for power. Despite very similar circumstances, Wilson had decided not to help the U.S. intervene, thus indicating that Blair's preference for greater control over world affairs may have been the key factor underlying his different course of action.

In sum, there is converging evidence from a variety of researchers in the literature that those leaders or governments who have a concern for having impact or prestige are more satisfied by

implementing strategies of force, or at least are intolerant of situations requiring patience, compromise and careful positioning. High power motivation therefore could be a critical indicator for predicting genuine adversarial intent in others.

However, this does not mean that power motivation would be sufficient. Winter (1987b) has also discovered, for example, a significant relationship between the use of power imagery by U.S. Presidents and instances of war avoidance. It is important to remember therefore that whilst the relationship between power motivation and the escalation of conflict may be consistent, it remains indirect. Dominating strategies that emanate from those high in their concern for power can lead opponents who are bluffing to back away. For this reason, one cannot assume that all instances of high power motivation in political leaders will lead to aggression. This, in turn, provides all the more reason to look for other prominent psychological indicators of aggressive intent.

1.2.2 Complexity

Usefully, a concern for power is not by any means the only psychological factor that is related to the use of aggressive or co-operative strategies. A wealth of research by Peter Suedfeld (for summary, Suedfeld, Guttieri & Tetlock; 2006) and his colleagues has shown that a leader's information processing style, as measured through their *integrative complexity* (referred to more generally as 'complexity'; Baker-Brown et al., 1992) and defined as their tendency to perceive concepts as both differentiated (separate and independent) and integrated (related to each other), can be just as important a factor in indicating their preferences for dealing with crises. Those leaders or decision-makers who have a reduced ability and/or reduced disposition to think about an issue from more than a single perspective tend towards simpler strategies of tough action, whilst those who hesitate to make simplistic all-or-nothing conclusions and have a high tolerance

for ambiguity are much less likely to resort to violence. They, in contrast, are more likely to prefer more nuanced ‘diplomatic’ routes to dealing with conflicts of interest.

There is now an accumulation of evidence to show that the complexity of political statements tends to decrease as a crisis reaches its point of escalation into conflict, whilst complexity levels surge prior to cessations of violence and peaceful resolutions. For example, in an early study demonstrating this phenomenon, Suedfeld and Tetlock (1977) examined the lead up period to two international confrontations that ended quite differently – the 1914 crisis in Europe that resulted in World War I and the Cuban Missile crisis in 1962 between the US and the Soviet Union. The complexity analyses in each crisis revealed a strong pattern. Communications from the crisis in 1914 were found to exhibit lower complexity in total over both phases (both in the preliminary and climax phases of the crisis) compared to the Missile Crisis in 1962. Moreover, the 1914 crisis communications exhibited further *drops* in levels of complexity when entering the climax phase (i.e., the time period just preceding the war) whilst at the equivalent time point in 1962, communications *increased* in their levels of complexity.

More recent studies by Winter (2007, for multiple crises between multiple actor pairs) and Walker and Watson (1994, contrasting the Munich and Poland crises between Britain and Germany) have used similar research designs and have replicated these findings. The presence of lower absolute levels of complexity before conflict is therefore a reliable indicator of crisis severity.

In addition to single outbreaks of large-scale conflict, complexity analyses have also been applied to protracted crises – for example, the crisis between the Israelis and their Arab neighbours, Egypt and Syria (1947-1976) (Suedfeld, Tetlock & Ramirez, 1977), and the crisis over the control of Berlin (1946-1962) between the US and USSR (Raphael, 1982). Here,

researchers found that the varying complexity scores of the protagonists' communications were indicative of the predicaments faced by the respective actors involved in each crisis. Political speeches preceding outbreaks of armed conflict (Middle East) or high tension (Berlin) were persistently lower in complexity than those that did not precede these developments.

This pattern of complexity pre-empting changes in political strategy over time has been replicated many times. Wallace, Suedfeld and Thachuk (1993), for example, have replicated this finding for the 1990 crisis in the Gulf, observing a significant drop below baseline complexity before Iraq's attack on Kuwait in 1990, whilst Tetlock (1985) has made similar observations regarding relations between the US and USSR (1945-1983), finding that more co-operative foreign policy solutions came with higher levels of complexity and that competitive (confrontational) policy came with lower complexity. Since these studies, Guttieri, Wallace and Suedfeld (1995) have made identical findings for the utterances of US decision-makers during the Cuban Missile crisis whilst Koo, Han and Kim (2002) did the same for the communications of South and North Korean policymakers over the years 1984 to 1994 – a period consisting of distinct phases of tension building and co-operation between the two countries.

Interestingly, integrative complexity is not the only measure of complexity that has been found to relate to fluctuations in aggressive/cooperative strategies over time. Hermann's (1980) measure of complexity, *conceptual complexity*¹, based on the ratio of complex (e.g., 'maybe', 'perhaps') to simplistic (e.g., 'always', 'absolutely') words used by a leader has been found to be moderately related to the political friendliness of the leader, a finding also made for forty-one U.S. presidents using integrative complexity (Thoemmes & Conway, 2007). Using yet another

¹ It is important that the term 'conceptual complexity' is not confused with the precursor variable to integrative complexity developed by Schroder, Driver and Streufert (1967). Hermann's conception of conceptual complexity is based on word counting, whilst Schroder, Driver and Streufert's method is essentially the same as integrative complexity, though they conceptualised complexity to be a mental 'trait' more than a mental 'state'.

measure of complexity², *cognitive complexity*, Maoz and Astorino (1992) have studied the structure of Anwar Sadat's thinking during a tumultuous period of time (1970-1978) between Egypt and Israel. Using a time-series regression analysis, the authors found that Sadat's cognitive complexity estimates predicted the number of co-operative Egyptian acts that came in the following time periods, with greater complexity associated with greater subsequent co-operation.

This latter finding followed earlier work linking cognitive complexity with pro-war and pro-peace arguments made by Israeli prime ministers. Maoz and Shayer (1987) found, in their analysis of twenty-one speeches made by four prime ministers in the Knesset, that justifications made for peaceful policies (e.g., the withdrawal from Sinai in 1957) were inherently more cognitively complex than justifications made for aggressive resolutions (e.g., the Six Day War).

In addition to having a reliable relationship with aggression, complexity has been shown to have another useful property - it provides a forewarning for outbreaks of conflict, even when these outbreaks are unexpected. Suedfeld and Bluck (1988) found in a study of nine surprise attacks that the average complexity levels observed in the communications of the attacking side evidenced a drop prior to the mobilisation of its military power, despite the absence of any explicit sign of intent. Moreover, this trend was consistent in eight out of the nine attacks³, indicating that it was extremely reliable. This evidence suggests that integrative complexity analyses can disclose subtle psychological shifts that are not otherwise obvious.

² Their measure of complexity is derived from a cognitive mapping technique developed by Axelrod (1973; 1976). Relationships between concepts are derived from text and represented in a matrix, from which different complexity calculations (four in total) can then be derived. It contrasts considerably with the measures used in this investigation (see Method Section for the other methods).

³ The exception was the Argentinean attack which triggered the Falklands War.

Another useful property of complexity is that it need not necessarily be measured in the communications of governments to be indicative of conflict escalation. For example, Suedfeld (1992a) studied the integrative complexity of newspaper editorials (from the Toronto Globe and Mail, NY Times and Pravda) in relation to the nature of bilateral events occurring between different country dyads⁴. Downward shifts in complexity found in these national newspapers were found to concur with negative events affecting the respective country pairs. Given that the media was distant from the political decision-makers, it was remarkable that any such relationship could be found.

Lastly, it appears that integrative complexity, even when measured on a day to day basis, can be highly sensitive to changes in the thinking of political actors. This has been demonstrated by Liht, Suedfeld and Krawczyk (2005), who examined negotiations between Chiapas guerrillas and the Mexican government in 1995 by analysing complexity levels measured from the verbatim transcriptions of the discussions. As with previous studies, they found that independently reported advances in negotiations were positively related to increases in daily levels of integrative complexity, thus signifying the willingness of both sides to search for a resolution. It seems clear then that leader intent (in this case, negotiator intent) in high pressure situations can change as quickly a leader's complexity changes, making this an ideal variable with which to examine leader intent during the pivotal/fast-moving stages of a crisis.

In sum, this wide variety of evidence on complexity points to the same underlying principle with regard to behaviour during political crises. Shifts towards lower levels of complexity seem to be a precursor of more implacable and aggressive policy stances such as the mobilisation of military

⁴ The country dyads were made up of the USSR, US, Canada and China respectively (though no editorials from China were used). Editorials were used when they were written chiefly about the other countries. Events occurred between the years 1947 to 1982.

force. Meanwhile, shifts to higher levels of complexity precede more open-minded, peace-seeking policy shifts that are likely to lead to compromise than war. Thus, like power motivation, these studies together show that complexity shifts within an individual over time, apparently even when measured by quite disparate methods, is a reliable indicator of policy behaviour.

However, as with power motivation, there are also some exceptions with regard to complexity findings. Levi and Tetlock (1980) for instance found in their examination of the Japanese attack on Pearl Harbour that integrative complexity measurements, taken from Japanese internal documents, revealed no discernable decreases in complexity levels (compared to baseline levels) just prior to the attack⁵. Meanwhile, in Maoz and Astorino's (1992) study of Egyptian-Israeli relations it was found (in contrast to Sadat's speeches) that Israeli prime ministers' speeches did not predict subsequent acts of co-operation on Israel's part. The question of why complexity did not predict the Japanese attack or Israeli co-operation in these instances remains an unanswered question and provides an important reason to temper a potentially excessive claim of a perfect predictive relationship with aggression/cooperation.

1.2.3 Optimism/belief in ability to control events

Though power motivation and complexity are the two most prominent indicators of adversarial intent in the literature, there is also some evidence accumulating for the importance of other psychological factors. Two such factors are the decision-makers' optimism (measured by their propensity to attribute positive events to internal, global and stable causes), and their belief their ability to control events (measured by the degree to which they attribute the locus of control in actions to themselves, regardless of whether these actions are positive or negative). Though the

⁵ They found instead that complexity levels were chiefly dependent upon to whom communications had been directed. When plans were being formulated in private, communications were less complex than when they were presented to the Emperor for approval. Differences between private and public communications are a concern addressed by Marfleet (2000, see Discussion section) but thankfully were not an issue in this study.

relationship between these variables and adversarial intent may not be immediately obvious, it is clear that before a leader can act in way that involves risk (such as launching an attack), they must have some degree of confidence in how they believe events will unfold, or that they have some control over the outcomes.

This relationship between optimism/control and political aggression has been shown in three studies. In the first, Satterfield and Seligman (1994) examined optimism and pessimism from spontaneous comments made by George H. W. Bush and Saddam Hussein prior to a number of their foreign policy decisions - decisions which varied in their risk and aggressiveness. They established that more pessimistic statements by either leader were more likely to be followed by cautious foreign policy decisions (e.g., withdrawing troops, making concessions) whilst an increase in the level of optimism in public statements preceded more aggressive and risk-laden strategies (e.g., attacking one's opponent).

Later, Satterfield (1998) tested this relationship even further, this time by combining integrative complexity with optimism/pessimism to predict aggressive and risky foreign policy decisions. Four new political leaders were examined – Adolf Hitler, Josef Stalin, Franklin Roosevelt and Winston Churchill - as before, in relation to the aggression in their decision-making. The author found that up to 36% of variance in the aggression in decision-making was explainable by the synergistic interaction between the leaders' complexity and optimism, suggesting that a leader's level of optimism may be a factor which enhances the relationship between low complexity and aggression.

Lastly, Dyson (2006) has found that a belief in ability to control events may have also been a key factor in Tony Blair's decision to enter the 2003 Iraq (along with his higher than average 'need

for power', see above), thus providing one final piece of evidence pointing towards the importance of aggressive decision-makers having belief in the course of action they wish to take.

1.2.4 Beliefs about the political world/distrust

Another factor that has been emerging as a significant indicator of adversarial intent is a leader's beliefs about the political universe (i.e., whether they see the world as aggressive or co-operative and whether they are distrustful of political others). Since Walker, Schafer and Young (1998; Young, 2000) defined a quantitative method of studying a leader's operational code (the set of beliefs a political leader has about the political world, as developed by Leites, 1951; George, 1969; Holsti, 1970), a number of studies have emerged that relate leaders' beliefs to their political behaviour. These have established that decision-makers' philosophical views of the political world are consistent indicators of their preferences for aggressive versus co-operative action when dealing with political others.

For example, a study of foreign ministers and secretaries of state by Crichlow (2005) found that when foreign policy preferences of the government were ranked on a continuous scale of aggression/co-operation, the level of the government's aggression was predictable from the foreign minister or secretary of state's attitudes about the political world. Specifically, he found that when officials made a large number of negative statements about political others (e.g., other countries, other leaders) or made statements that illustrated a great distrust of those political others (see Driver, 1977; Hermann, 2006; for measurement), those same officials also scored higher on policy aggressiveness.

Building on this earlier work, Schafer and Walker (2006) found evidence linking beliefs about the political universe and variability in the so-called *democratic peace* - the phenomenon that leaders of democracies choose not to go to war with one another, whilst conflicts between democracies and non-democracies remain more likely. They did this by examining the

operational codes of two democratic leaders, former Prime Minister Tony Blair and former President Bill Clinton, during the Kosovo crisis. Whilst the authors found that both Blair and Clinton had more co-operative beliefs about democracies than non-democracies, they also uncovered subtle differences in the way Tony Blair and Bill Clinton treated non-democracies. Beyond the two leaders' negative beliefs about non-democracies, the authors found that Tony Blair was verbally even more hostile to non-democracies than Bill Clinton. Moreover, this difference seemed to manifest itself in the actual policy behaviour of each leader's country during the Kosovo crisis. Whilst both the U.K and the U.S used equally co-operative strategies towards *democracies* during this crisis, the U.K. was significantly more hostile than the U.S. towards *non-democracies*.

Shannon and Keller (2007), in a study of the recent Bush administration's attitudes towards Iraq policy, also had results that support this trend. Across all major administration members, the variable that was most associated an official's desire to break international norms and invade Iraq was distrust. This seemed to be systematically greater amongst the Bush administration's Iraq 'hawks' (e.g., Bush and Paul Wolfowitz) than amongst those who were more reluctant to use military action (e.g., Colin Powell and Richard Armitage).

An additional property of beliefs about political others is that they seem to be shaped in response to events. Work by Walker, Schafer and Young (1998) has shown that Jimmy Carter's beliefs about the political universe changed considerably in response to the Soviet invasion of Afghanistan. Prior to this event Carter had for three years maintained a positive, co-operative and optimistic view of international relations as reflected in his positive evaluations of political others. However, these shifted significantly in the negative direction after the invasion, indicating that beliefs are malleable in response to changing circumstances.

This finding has been consolidated by other research showing belief shifts in important figures. Crichlow (1998), studying the operational codes of Shimon Peres and Yitzhak Rabin, found that both these figures developed a more positive view of the world over time. Peres and Rabin began in the 1970s with a very negative view of the political universe - a universe that was hostile to Israel. But this world view softened considerably over the next twenty years so that, by the 1990s, each figure had reverted to an essentially neutral world view. This significant change in the leaders' perceptions coincided with Israel's increasing acceptance by previously hostile nations such as Egypt and Jordan.

The fact that leaders' beliefs about the political world seem to be strong indicators of co-operative versus aggressive policy preferences in politicians makes these variables potentially very useful for detecting shifts in leaders' intentions. Though there is not as much evidence linking beliefs to policy behaviour as, for example, integrative complexity or power motivation, the quantitative versions of operational code variables have not been available for nearly as long. There is therefore great potential for more work demonstrating these relationships and, in the mean time, they should certainly be considered when formulating any predictive model of adversarial political action.

1.2.5 Summary of psychological indicators of intent

Overall, this review has summarised a number of studies which show that psychological indicators could be applied to predict aggressive/cooperative political actions. When leaders or decision-makers in a political crisis have a cynical view of the political world, are concerned about maintaining their control or status, see the world and its dimensions as 'black and white' and are optimistic about their ability to produce their desired outcomes, this seems to be a psychological recipe for escalation into conflict. In contrast, when leaders in a political crisis see the world as co-operative, are relatively unconcerned with being dominant, see the world as

nanced and multi-faceted, and are pessimistic about their ability to attain desirable outcomes, this leads to the use of more diplomatic strategies. Though none of these factors is likely to provide perfect predictions on their own, together they offer potentially enormous predictive power if applied to crisis behaviour. To date, this predictive power has never been fully investigated.

1.3 Applying Psychological Factors to Predict Aggression/Cooperation

Now that potentially useful predictors during political crises have been extracted from the literature, it is possible to turn to the question of how to apply that potential predictive power. How can these variables be best employed to make sound predictions about whether crisis behaviour is about to take a violent (or placatory) turn?

As stated at the beginning of this introduction, most political psychological work has been more concerned with explaining events and so approaches to applied prediction have not been addressed in detail. At the same time however, this literature does suggest two broadly different frameworks for making applied predictions - the 'profiling' and 'tracking' approaches respectively - by the way researchers have approached their explanation of events. Whilst profiling has actually been used to make predictions about the actions of political leaders (see below), no-one has as yet fully addressed the applied use of a tracking approach. As will be explained, though the profiling approach to prediction has already been partially established, there are strong reasons to see advantages in the alternative tracking approach – a contention that is tested empirically here.

1.3.1 Profiling approach

The 'profiling' or 'individual differences' approach to explaining political behaviour models leaders' intrinsic qualities or *traits* as being one factor of many that influences the determination

of foreign policy. Though there is a degree of variety with regard to the use of profiling techniques, the common assumption underlying all profiling approaches is that individuals bring different qualities to the decision-making process and that these qualities can affect problem interpretation and solution preference (Hermann & Hagan, 1998, Hermann et al., 2001). Thus, in order to predict what decisions a leader is likely to make, his or her qualities must first be measured and then made meaningful through the development of a leader profile.

To create such a profile, there is a general procedure (developed chiefly by Hermann, 1980, 2006) to which other researchers have implicitly subscribed (e.g., Winter, 2001). This involves:

- i) Firstly measuring aspects of leaders' responses to questions at various time points⁶ (using variables such as those described in the first part of my introduction)
- ii) Secondly, comparing these measurements to a number of leaders (thereby making them meaningful)
- iii) Thirdly, using these leaders' relative scores to determine their profile *or* political leadership orientation (e.g., Constraint-Challenging or Active-Independent, Hermann, 2006).

The major assumption behind creating such profiles is that those with the same profiles will also *behave* similarly and, as one might expect, there are studies showing that leadership profiles correlate with different behavioural tendencies.

For example, in a much cited investigation of forty-five different leaders, Hermann (1980) found that leaders who have an *independent* orientation (i.e., a combination of high distrust in others,

⁶ Usually, but not always. Winter's (2001, personal correspondence) profiles on George W. Bush and Barack Obama are based on comparisons with other Presidents principally at the moment of their inauguration ceremonies – just one time point.

high need for power, low complexity, and a strong belief in the ability to control events) were more likely to receive negative feedback from other countries and to pursue hostile policies. On the other hand, those leaders with *participatory* orientations (i.e., broadly the opposite characteristics to those with independent orientations) had the opposite relationship with these outcomes.

However, despite evidence for patterns across many different individuals, it is a very different task to use profiles to make predictions for a specific individual – the kind of predictions that one would want to make when predicting action during a political crisis. This can be illustrated by reviewing two attempts to actually predict the behaviour of political leaders by forming profiles of them in real time. These studies attack the concerns of this project most directly, so it is certainly worth dwelling on them.

The first attempt by Winter, Hermann, Weintraub, and Walker (1991a) was a joint project that sought to profile Mikhail Gorbachev and George H. W. Bush, predicting how each leader would manage their respective leadership roles. Each author in the project used their own idiosyncratic profiling approach to make their own predictions. Here, Winter and Hermann's profiles are of particular interest, given their use of some the variables reviewed in the first part of this introduction.

For example, using her leadership trait analysis approach (Hermann, 2006, 1987, 1980), Hermann classified Gorbachev as having a *developmental* orientation - high in, for example, distrust and complexity. She predicted that Gorbachev would pursue a policy of 'controlled independence', that he would create opportunities through co-operation with others and would do this without over-committing resources. Using the same criteria, Bush was classified as

having an *integrator* orientation. He was predicted to be interested in improving the standing of the country and would do this by gaining a broad basis of support.

Exploring the motives of each leader, Winter found that Bush scored well above average on the achievement (i.e., one's desire for excellence) and affiliation dimensions (i.e., one's desire to have social involvement and contact with others) but was average on the dimension of power, with Gorbachev being essentially the same. Due to their high scores on the affiliation and achievement dimensions, Winter predicted that both leaders would be rational and co-operative, would seek arms limitations (as has been previously indicated by high affiliation imagery; Winter, 1987b) and would use aggressive tactics sparingly providing neither leader was betrayed. In circumstances where either leader was betrayed by the other, however, they were predicted to be aggressive in their response (as is the case with those high in affiliation motivation).

In a follow-up paper, the same authors (Winter, Hermann, Weintraub & Walker, 1991b) evaluated these profiles for their insight and accuracy. They claimed that the general predictions that had been made were given support by the subsequent behaviour of the two leaders, especially with respect to two major events: the Persian Gulf crisis and the sliding economic and political circumstances in the USSR.

For instance, with regard to Bush's behaviour during the Gulf crisis, one might say that his coalition-building prior to war was evidence of his desire to build support before acting (predicted by Hermann). This contrasts with Bush's son, George W. Bush, who was less patient in forming a broad base of international support before the 2003 Iraq War, as is consistent with someone relatively high in power motivation⁷. Bush's decision to use force in response to

⁷ Note that George H. W. Bush was only moderate in his power motivation (Winter et al., 1991a)

Hussein's invasion of Kuwait also fell in line with Bush's high affiliation motivation (as predicted by Winter). Bush had been a supporter of Hussein but his unilateral infringement of Kuwait's sovereignty brought out his reactionary side – a common response from someone who is high on this dimension and is let down by an ally.

Meanwhile, Gorbachev's desire to keep the Soviet Union together was indicative of his overall developmental orientation, whilst his failure to achieve his policy ambitions was also a result of a lack of political muscle on his part that comes with a merely average degree of power motivation. Gorbachev's pragmatism also expressed itself in his handling of the Persian Gulf crisis, promoting a variety of strategies with respect to Iraq, including the use of incentives – in turn consistent with his higher than average conceptual complexity.

Given these occurrences, one might conclude that both Gorbachev and Bush were well-characterised by their profiles. But there is a counter argument that one should not take such evidence of success wholeheartedly. The predictions made by the researchers were broad, providing only an impression of what to expect. Unfortunately, making apparently convincing connections between real-life events and predictions, especially when the predictions are not especially specific, is a common human tendency. The *Barnum effect* - demonstrated for example when people attribute specific elements of their behaviour as having been captured accurately by vague personality profiles - could be occurring here in the authors' post-hoc interpretations of these quite general predictions (see Furnham & Schofield, 1987). One could, for example, argue that George H. W. Bush's decision to enter the Gulf War against Iraq was completely inconsistent with his merely average level of power motivation compared to other leaders (a point that Winter et al., 1991b, do in fact acknowledge) rather than arguing that the aggressive response was due to his high affiliation motivation. The problem of course is that whether or not Bush went to war against Iraq, at least one prediction had to be correct, thereby

making the prediction *half* right. Either Bush went to war because of interaction between Hussein's betrayal and his affiliation motivation, or he didn't because of his lower power motivation.

More generally, the problem with deriving predictions from profiles in advance of context is that they have to be general in their nature. As Winter et al. (1991b) noted, they could not have foreseen the invasion of Kuwait by Saddam Hussein's forces and so could not address the likely reaction of Bush to this development. They could only address hypothetical possibilities. This in turn gives rise to the problem of falsifiability. If only a general impression can be formulated as a prediction, how can that same prediction then be deemed true or untrue? Winter et al.'s predictions may in fact have been sound, but without *specific* predictions, it is very difficult to evaluate them tangibly.

A later prediction by Winter (2001) for George W. Bush does seem to do better in this regard. Based on Bush's high power motivation, high affiliation motivation, and low achievement motivation, Winter (2001) projected that Bush would be aggressive in his foreign policy with regard to rogue states such as Iraq (based on his high power motivation) if he listened to some of his "hawk" advisors, that he would enjoy being president (based on his high power and low achievement aspirations) and that he would be vulnerable to scandal due to the closed and secret nature of his decision-making, and to the strong influence of advisors on his foreign policy (in turn due to his high affiliation motivation).

At first glance, Winter's (2001) predictions seemed to have improved on earlier predictions by being more specific. For example, he articulated that Bush might be more aggressive in his policy towards Iraq – a prediction that was dramatically confirmed when the U.S. invaded Iraq in 2003. Bush's presidency was also littered with scandal. The controversial unveiling of

undercover CIA agent Valerie Plame for political reasons is thought by one Bush administration insider to have been due to the influence of Bush's advisor Karl Rove (McClellan, 2008), something to which Winter predicted Bush would be susceptible because of his high affiliation motivation. Even though many Presidents' administrations have moments of scandal (whether it is their fault or not), Winter clearly identified that in Bush's case this would be due to the influence of advisors and not, for example, due to hedonistic tendencies (e.g., as with Bill Clinton).

However, what is not entirely clear is how much of these predictions emanated from the use of psychological variables and how much was derived from Winter's *general* expertise about the Bush administration. For instance, the prediction was clearly boosted greatly by Winter's (2001) awareness of Bush's advisors (in particular, the 'hawks', whom Winter pointed out specifically) who would ultimately have great influence on Bush's foreign policy. Could Winter perhaps have made the prediction about Iraq in the absence of a profile?

A recent profile by Winter (personal communication, April 24, 2009) of the new U.S. President Barack Obama adds to these concerns. In this profile, Obama is shown to have a very similar profile to George W. Bush, who in turn has a similar profile to John F Kennedy. In particular, all three are well above average in their measured degree of power motivation when compared to other U.S. presidents. The reason this is a problem, of course, is because it is known that Kennedy did not choose an aggressive response to the Cuban Missile Crisis and, even more relevantly, Obama did not support the 2003 invasion of Iraq. What then can explain these inconsistencies?

Perhaps the most probable explanation is that the motives of these leaders changed dramatically over time. Obama and Kennedy's power motivation may have been similar to Bush's at their

inauguration speeches but may have been different in their reaction to the Cuban Missile Crisis and the build-up to the 2003 Iraq War. This is partly supported by Winter's own data on Barack Obama, which show that prior to being elected, Obama was decidedly average in his degree of power motivation compared to other U.S. Presidents (Winter, personal communication, April 24, 2009). His power motivation only increased to Bush's level after he was elected.

Does this mean that Obama would have supported the Iraq War had he been President in Bush's time, given the high power motivation in his inaugural speech? Assuming that such a war would have been suggested by Obama's advisors, it would probably still have depended on Obama's *reaction* to such a suggestion - just as Obama's higher power at his inauguration speech may have been a reaction to the economic situation at that point (i.e., his perception that the nation needed to see a strong leader during the economic crisis) or perhaps a reaction to the inauguration itself - a moment of empowerment and prestige. It is also even possible that Obama would have been aggressive in certain aspects of his policies (e.g., Afghanistan) but not in others (e.g., Iraq). After all, high dispositional power motivation is not deterministic of everything, nor do researchers claim it to be. In sum, the point is that one would need to see Obama's power motivation in all those different contexts to answer these questions with any certainty.

In conclusion, Winter's (2001) prediction for George W. Bush may indicate Winter's expertise about political leaders, but it does not necessarily promote a profiling approach to prediction any more than the profiles of Bush senior and Gorbachev. Moreover, the fact that leaders can and do change over time and in reaction to different events suggests that a predictive approach that is (i) sensitive to changes over time, that is (ii) data-driven and that (iii) does not involve a large degree of subjectivity in its formation and validation, would be preferred. It is to such an alternative approach that this introduction now turns.

1.3.2 Tracking approach

In view of the problems associated with a profiling approach to prediction, it is useful to explore another possibility suggested implicitly in the literature – an approach that shall henceforth be called the *tracking* approach to prediction. A tracking approach to prediction makes use of an entirely different property of psychological indicators to the profiling approach. Rather than making predictions based on individual differences, this approach makes predictions by monitoring psychological indicators as they change (or do not change) over time. Psychological shifts, it is contended, precede shifts in political strategy.

The first part of this introduction touched on many examples of how this might work – albeit hypothetically. For example, retrospective analyses of integrative complexity over time have shown that complexity levels drop in the communications of the attacker country shortly before the attack occurs (Suedfeld & Bluck, 1988). In theory, monitoring complexity over time (along with other indicators) could provide substantial predictive power when predicting sudden upsurges in the use of violent or dominating political strategies *or* major shifts toward co-operation.

Monitoring indicators over time could also bestow a number of advantages over the profiling approach. The more data-driven process of monitoring leaders means that analyses would always be up-to-date and so interpretable in the current political context. As the political situation changed, measurements of psychological characteristics could be used to reflect the decision-makers' response to this change, whether their information processing and motivations are affected, and how confident they would be of taking risky choices. Few assumptions would need to be made about the future situations a leader or government would face. This contrasts favourably with the profiling approach where, once a profile is constructed, there is little contextual information immediately available to guide predictions - the situations that a leader

could face must also be hypothesised. Few assumptions of this kind are required in a tracking approach of course.

Though the features of a tracking approach contrast favourably with a profiling approach, there is a dearth of literature exploring its actual application. This is somewhat surprising given that Suedfeld and Rank (1976) were suggesting, over thirty years ago, that researchers might make real-time predictions with this approach and let the course of history confirm or disconfirm their hypotheses.

1.4 A Study Testing the Utility of the Tracking Approach to Prediction

The following empirical case study was therefore designed to address this gap in the literature by applying several psychological indicators of adversarial intent to the task of predicting the aggression and co-operation levels of a political actor during a crisis – in this instance, the Government of Zimbabwe. In order to do this, the power motive imagery (Winter, 1973; Winter, 1991c), need for power (Hermann, 1980; 2006; Winter, 1973), integrative complexity (Baker-Brown et al., 1992), conceptual complexity (Hermann, 1980; 2006), belief in ability to control events (Hermann, 1980; 2006) and beliefs about the political universe (Walker, Schafer & Young, 1998) of the Zimbabwe Government (as measured through its propaganda) were measured over the course of 2007-2008 with the express purpose of (i) fitting a stochastic model to the data and (ii) using this model to make predictions in 2009. At the same time, the frequencies of violent and co-operative acts committed by the Zimbabwe Government against/with political opposition were independently measured as indicators of the Zimbabwe Government's changing crisis strategy.

In sum, this research was designed to evaluate a tracking approach to prediction by providing answers the following five questions:

- 1) Are these psychological indicators of intent related to aggressive and co-operative acts by the Zimbabwe Government, as previous research has shown?
- 2) In addition to being related to violence or co-operation, do psychological indicators predict violent and co-operative acts?
- 3) If the psychological indicators are predictive, how much variance in political strategy can be predicted from a tracking approach to prediction (i.e., just how effective is the predictive utility)?
- 4) Over what time-frame are psychological predictions of violence possible and/or optimally made?
- 5) Does a model derived from the historical data extrapolate to predictions made in real time, or alternatively, 'out of sample'?

The findings of previous research allow strong predictions to be made for the answer to the first question and give some guidance for the second question. For the rest of the questions, however, making any specific prediction was difficult because no consistent findings have been reported in the literature.

For question one it was predicted, based on the research already reviewed, that greater violence would be related to lower complexity (both integrative and conceptual), greater need for power or power motive imagery, greater belief in ability to control events and a belief that the world is not a cooperative place. For question two, it could also be predicted that changes in power motivation and integrative complexity would both likely *precede* changes in the use of violent versus co-operative strategies (e.g., based on Winter 2007; Winter, 2003a; Suedfeld & Bluck, 1988), though it was less clear whether the other variables would predict violence/co-operation.

For the last three questions, there was very little guidance for any specific expectation. For example, only a few studies (e.g., Satterfield, 1998; Maoz & Astorino, 1992) have explicitly measured the amount of variability in behaviour explained as a function of predictors (such as in a regression analysis). This was not enough to specify an expectation. Meanwhile, there was little from previous research to speak to question four, given that studies have examined indicators over time at various intervals (e.g., Winter, 1993a, Liht, Suedfeld, & Krawczyk, 2005; Suedfeld & Bluck, 1988, Maoz & Astorino, 1992), with no clear emerging norm.

Lastly, the potential for a predictive psychological model of adversarial/co-operative intent to be predictive out-of-sample (i.e., used to make predictions outside of the sample used to derive the model) was also entirely unknown, given that no researchers have yet tried to develop such a model for this purpose.

Overall, it was expected that ultimately two time-series regression models would be fitted to the data and tested out of sample during the course of 2009, though the variables used in the model (other than power motive imagery and integrative complexity) and the unit of time aggregation ('t') were uncertain and would have to be determined through exploratory analyses. To represent this uncertainty, variable 'X' was allowed to be one (or more) of the other aforementioned variables.

These hypothesised regression models could be written formally as follows⁸:

Model 1

$$\text{Log(Violence}_t) = \beta_1 \text{Power Motive Imagery}_{t-1} + \beta_2 \text{Integrative Complexity}_{t-1} + \beta_3 \text{Variable } X_{t-1} + \beta_0$$

⁸ The dependent variables in these equations are log-transformed because this is the optimal link function for count data when the data is Poisson-distributed. More regarding this can be found in the Results section.

Model 2

Log(Cooperation_t) =

$$\beta_1 \text{Power Motive Imagery}_{t-1} + \beta_2 \text{Integrative Complexity}_{t-1} + \beta_3 \text{Variable } X_{t-1} + \beta_0$$

2 METHOD

2.1 Selection of a Crisis Test Case and Scoring Materials

In order to provide a fair trial for the tracking approach to predicting conflict/co-operation, it was necessary to select a political crisis test case that had two distinct features. Firstly, the crisis had to have lasted over time, and therefore provide a long enough time period (or sample size) from which a statistical model could be appropriately fitted. Secondly, the crisis had to have given rise to periods of violence but also periods of relative peace. This was so that psychological indicators could be examined over the full range of strategies being used by the political actor in question (i.e., government, group or leader) and used to fit an accurate stochastic model that could subsequently predict the same range of strategies in real-time⁹. A political crisis that fulfilled both of these criteria was the crisis (still currently ongoing) in Zimbabwe.

Zimbabwe has been experiencing a long-term political crisis since the year 2000 when the Zimbabwean Government, led by Robert Mugabe's Zimbabwe African National Union - Patriotic Front (ZANU-PF) party, began its compulsory takeover of farmland from commercial farmers. Since then, the economy has declined substantially, whilst Robert Mugabe and ZANU-PF have been struggling to hold onto power in the face of political opposition from the Movement for Democratic Change (MDC), led by Morgan Tsvangirai. During this period of internal political struggle, ZANU-PF has shown a willingness to combine brutal violence (e.g., the beating of Morgan Tsvangirai in March 2007) with acts of co-operation (e.g., the signing of a power-sharing government in September 2008) in order to maintain power. This contrasting use

⁹ Range restriction, which occurs when variables are not measured over their full potential range, is a common and often ignored phenomenon leading to the underestimation of the relationship between variables (Glass & Hopkins, 1996). Given that the initial forecasting model for the case study was to be based on historical data, this meant that it was sensible to select a period of time when a full range of behaviour by a political actor had been observed.

of violence and co-operation by the Zimbabwean Government in dealing with its political opposition over the 2007/2008 period therefore makes it ideal for fitting a regression model of adversarial intent which could then be used to make real-time, data-driven predictions during 2009.

With a crisis case chosen, the issue arose of whom one should examine within the Zimbabwe Government as a proxy for its intentions. This proved a more difficult task than the selection of the crisis test case. Though Robert Mugabe (the long-time President of Zimbabwe) would have been the obvious choice for this purpose, a survey of the available materials revealed that over the course of the year 2008, only seven complete speech transcripts could be found. This was not enough material to measure psychological indicators across the span of that year. Disappointingly, even less could be obtained for other key leaders within ZANU-PF such as Vice-Presidents Joice Mujuru and Joseph Msika.

However, one excellent source of material was available – the editorials of Nathaniel Manheru, a writer for the Zimbabwe state-run newspaper *The Herald*. These articles were useful because the name ‘Nathaniel Manheru’ is actually a pseudonym for Robert Mugabe’s chief spokesperson, George Charamba. The newspaper column had originally been set up by the former information minister Jonathan Moyo as a means of surreptitiously bombarding Zimbabweans with ZANU-PF propaganda. However, after his ejection from ZANU-PF, Moyo became a vocal critic of the Zimbabwe Government, revealing that it was he who had originally created the column and that it was George Charamba who had since taken ownership of it after Moyo’s departure in 2005 (Moyo, 2008).

In contrast to Mugabe’s speeches, the Manheru articles were abundantly available. For 2007, 31 of Manheru’s articles were found whilst an additional 28 articles were obtained for 2008. The

vast majority of these were easily extracted from the news archive website www.allafrica.com whilst the rest were pulled from the *World News Connection* database (<http://wnc.dialog.com/>). These articles thus provided enough information for the measurement of psychological indicators associated with the Zimbabwe Government across the initial two-year time span of interest.

The articles were also ideal in a number of other respects. For instance, they had a consistent audience (i.e., *Herald* readers), they were all authored by the same person, and they all spoke to current political issues. These consistencies across the articles meant that many of the factors that might adversely affect the direct comparability of content analysis measures across time were automatically held constant (see Schafer, 2000; Walker, 2000 for a summary of many of these potential issues). Researchers usually have very little control over potential effects of audience and subject domain on thematic content analysis, especially when scorable material is in short supply, so this was certainly a bonus when scoring Manheru.

Hence, given Charamba's closeness to the ZANU-PF leader (Moyo has even gone as far to say that "Mugabe speaks through Charamba"; Moyo, 2008), along with the fact that the articles were written for the Zimbabwe Government's interests, these articles made an ideal proxy for the government's perceptions and mindset, as measured through the lens of appropriate psychological indicators.

2.2 Independent Variables

The written materials obtained for the Zimbabwe Government were scored on four distinct psychological dimensions which are thought to be related to the use of aggressive political strategies. These were complexity, power motivation, belief in ability to control events and beliefs about the political world. For the scoring of two of these dimensions - power motivation

and complexity - two different methods were used, bringing the total number of variables scored to six.

The purpose of including two measures of the same variable was to compare the different methods of measuring the constructs, two of which (need for power and conceptual complexity) can now be scored by computer software (Hermann, 2006). This was of additional interest because no published research has yet used these two forms of the variables in a way that they could be statistically compared, even though computer software offers a welcome alternative to scoring by hand.

2.2.1 Complexity - Integrative complexity

As stated, two methods of scoring complexity were used. The first of these, called integrative complexity (IC; Baker-Brown et al., 1992; Schroder, Driver, & Streufert, 1967; Suedfeld, Tetlock, & Streufert, 1992), scored here according to the guidelines laid out in Baker-Brown et al. (1992), is a widely-used measure of the degree to which a person's perception of a stimulus is differentiated (i.e., perceived from different perspectives) and integrated (i.e., where perceived dimensions are conceptually connected). IC is scored for each idea expressed by the communicator (usually at the level of the paragraph) on a scale of 1 to 7, though scores of above 5 were not found in this particular study.

Scores for IC are based on the structure, not the content, of the ideas being expressed. Thus, a paragraph, for example, that conveys an historical event as having multiple distinct causes (i.e., differentiation) and states that these causes interacted with each other to produce the event (i.e., integration) would be considered high in its level of IC and would earn a score of 5. Furthermore, this score would be assigned regardless of what the different causes identified were.

In the case where the paragraph only expresses one possible cause for an event, this is said to demonstrate neither differentiation nor integration and would earn a score of 1. For intermediate cases, where a communicator may identify two or more causes of an event (differentiation) but does not identify any conceptual connection between those causes, the paragraph is given a score of 3, whilst scores of 2 and 4 are reserved for more ambiguous cases falling between the three major scoring categories¹⁰.

Though IC was scored here at the level of the paragraph, aggregated scores for IC over a particular time period were derived by finding the mean average of the paragraph IC scores that fell within that period. In this study, at least 14 paragraphs were scored for IC in any aggregated time period occurring across the two years (see time aggregation procedure below). However, in the vast majority of cases far more than 14 were scored, allowing precise estimates of mean IC to be obtained.

2.2.2 Complexity - Conceptual complexity

Conceptual complexity¹¹ (CC) (Hermann, 2006), like IC, is designed to measure the complexity of a person's cognition. However, unlike integrative complexity, CC is measured as a function of the *content* of a person's language and additionally does not attempt to capture the aspects of complexity that involve the integration of concepts. In the profiling method used by Hermann (2006), conceptual complexity is conceptualised as a trait¹² but this does not mean that it does not change over time. For example, Dille and Young (2000) have shown that a CC score, just

¹⁰ Scores of 6 and 7 are reserved for the demonstration of additional levels of integration. However, given that these levels were not demonstrated in this context, the reader is directed to the integrative complexity scoring manual (Baker-Brown et al., 1992) should they wish to know more about these.

¹¹ The original term 'conceptual complexity' was used by Schroder, Driver and Streufert (1967), but this is not the same as the variable measured by Hermann (1980; 2006), which is based on word counting. The original was scored using a similar method to integrative complexity.

¹² The same applies to belief in ability to control events.

like an integrative complexity score (e.g., Suedfeld, Corteen and McCormick, 1986) is at once both a function of a person's trait and state complexity.

For CC, a person's language is scored as higher in complexity as it becomes enriched with words pre-defined as complex (e.g., approximately, perhaps) versus those words pre-defined as 'simple' (e.g., absolutely, irreversible; Hermann, 2006). No difference is defined between words within the 'complex' or 'simple' categories. A word is coded as either complex or simple but not in degrees of complexity or simplicity.

It is calculated as a proportion in the following way:

$$\frac{\text{number of complex words}}{\text{number of complex words} + \text{number of simple words}}$$

Thus, a larger number of complex words *relative* to simple words results in a higher complexity score.

As with need for power and belief in ability to control events (see below), modern CC scoring is accomplished very efficiently through the use of the Profiler Plus© software package (<http://www.socialscienceautomation.com>), developed by Michael Young in conjunction with Margaret Hermann (see Hermann, 2006; Young, 2000). Because of this efficient scoring system, all of the available material could be scored for this study in a very short time.

The text provided by Manheru's articles was rich in scorable material, allowing for accurate estimates of CC for any given time period for which material was available. This meant that even the least accurate estimate of CC for any time period in the dataset was based on a total of 117 identified simple/complex words.

2.2.3 Power motivation – Power motive imagery

Power motivation, like complexity, was also scored using two different methods. The first of these methods is called power motive imagery, which has been developed and validated by Winter (1973) and was scored in this study according to the running text guidelines laid out in Winter (1991c).

Power motive imagery measures the degree to which the communicator expresses a concern for impact on or control over others in the world, or a concern for their prestige or status relative to others. It does this by identifying images of such concerns in the communicator's language. Common instances of power imagery include interactions where a "forceful action" is taken against another or, for instance, where advice is given to another person when it had not been requested by the receiver (Winter, 1991c). An example of the former is easily demonstrated from text scored in this study (power image in italics):

"After all *a blow that falls on the belly of an enemy*, never mind from where, is always most welcome."

Though scored at the level of the sentence, power imagery is measured by the number of power images per 1000 words scored – a convention that the developers have apparently found more useful in characterising the degree of someone's power motivation than, for instance, images per 100 sentences (Winter, 1991c). It is also scored alongside two other motives (achievement and affiliation motive imagery) which are not reported here because they were not of primary interest in building a model of adversarial intent, even though they are still scored as part of the integrated running text procedure outlined by Winter (1991c). Should the reader wish to know more about the scoring conventions for the other two variables, they are directed to that same source. However, they are not relevant to this particular work.

2.2.4 Power motivation - Need for power

The idea of ‘need for power’ (nPower; Murray, 1938; Smith, Atkinson, McClelland, & Veroff, 1992; Hermann, 2006) is conceptually identical to power motivation (see definition under power motive imagery section; Winter, 1973) but is measured in a dramatically different way. Rather than focusing on the images of power formed in the language of the communicator, scoring for nPower focuses on the number of *verbs* that suggest a concern for having impact or control over others (denoted as P+) versus those that do not (denoted P-).

The computer software (Profiler Plus©) that makes the decision of whether or not a verb phrase denotes power, does so based on predefined criteria (i.e., a pre-defined list of verbs in particular contexts). For example, the phrase “If **pressed**, I am ready to give names” is scored as demonstrating a concern for power (P+), whilst the phrase “I notice there is fascination with how the material was obtained” is scored as not demonstrating power (P-)¹³. The nPower score is then calculated as the proportion of *all* verbs scored as P+:

$$\frac{(P+)}{(P+) + (P-)}$$

All the available material for Manheru was scored on this psychological dimension. However, reasonable estimates of a Manheru’s need for power could not be obtained for every time period for which there was material available. Scores of need for power based on a total of 14 or less verbs for a given period (i.e., when $(P+) + (P-) < 15$) for a given time period) were dropped from the analysis and marked as missing data. This was in order to help avoid wild and misleading measurements that could disrupt statistical inferences and the fitting of a model¹⁴.

¹³ Examples come direct from material scored in this study’s dataset.

¹⁴ Dropping data from any dataset is never ideal. However, neither is keeping data when it could be drastically misleading. Within the classic regression framework used in this study, there is no method by which one can

2.2.5 Belief in ability to control events

Belief in ability to control events (BACE) is a measure of the degree to which a person perceives outcomes as having been self-generated rather than caused by others (see Hermann, 2006). It is calculated simply as the number of phrases indicating perceived internal control (IntCon) (e.g., “we can, as Zimbabweans, address problems and challenges on our own”)¹⁵ as a proportion of those perceived as either externally controlled (ExtCon) (e.g., “Equally, we have seen attempts by bigger nations at destabilising world peace”) *or* internally controlled. Thus, calculation of BACE is similar to that of conceptual complexity:

$$\frac{\text{IntCon}}{\text{IntCon} + \text{ExtCon}}$$

As with all the Profiler Plus© scored materials, BACE can be scored efficiently and with perfect test-retest reliability, so all available material obtained was scored for this variable. However, like the ‘need for power’ variable scored by Profiler Plus©, a reasonable estimate of BACE could not always be obtained for every time period examined. Thus, scores of BACE based on a total of 14 or fewer observed attributions of internal or external control for a given period (i.e., when $\text{IntCon} + \text{ExtCon} < 15$ for a given time period) were dropped from the analysis and classified as missing data. This was again in order to help avoid misleading measurements that could harm statistical modelling.

represent the uncertainty with which a predictor variable value is estimated (Cohen, Cohen, Aiken & West, 2003). The only way to counter this problem therefore is to ensure that all predictor values for a predictor variable are at least reasonably well-estimated. Consequently, in this case study, at least 15 observations were considered necessary to provide an estimate for a *proportion* or *ratio* variable.

¹⁵ Examples taken from this study’s dataset.

2.2.6 Philosophical beliefs about the political others (P1)

Positive versus negative beliefs about political others, measured by the variable ‘P1’ in operational code analysis (Walker, Schafer & Young, 1998), are a gauge of the degree to which a communicator views the political world as one of conflict or co-operation, with higher scores indicating a more positive view of the political world.

The variable is coded by Profiler Plus© software using Walker, Schafer and Young’s (1998; 2006) Verbs in Context System (VICS). In this system, all utterances are first divided into two of their main components - subjects and verbs. This information is then used to define whether the action in the sentence is done by ‘self’ (where the subject of the sentence is the person communicating, or someone with whom that person identifies) or ‘other’ (where the subject is anything other than that classified as ‘self’) and is either negative or positive (derived from the valence of the verb which is predefined prior to coding). In the specific case of Nathaniel Manheru, subjects that were defined as ‘self’ included Manheru himself, ZANU-PF, ZANU-PF officials such as Robert Mugabe and anyone affiliated with the party (e.g., the armed forces and the police). All other subjects were defined as ‘other’.

For calculating P1, instances of verbs with positive valences are first coded as ‘1’, while verbs with negative valences are coded as ‘-1’. These instances are then averaged for all utterances over the time period in question where the subject of the utterances has been classified as *other*¹⁶. This therefore renders a mean score between -1 (totally negative) and +1 (totally positive) as a measure of the communicator’s view of political *others* during that period.

Unlike other Profiler Plus©-scored variables, some additional manual modifications of the scoring were also required. For instance, because Profiler Plus© scored some subject categories

¹⁶ Those classified as ‘self’ were removed from the dataset.

as ‘other’ when they should have been scored as ‘self’ (e.g., when the subject was the Zimbabwean Government, Robert Mugabe, or ZANU-PF, and thus affiliated with Manheru) this had to be manually corrected by removing these utterances. However, given that the output from Profiler Plus© provides a list of the identified subject in each utterance scored, this adjustment could be achieved easily. With these utterances removed from the dataset, P1 could still be estimated accurately in every time period for which there was material available. The least accurate estimate of P1 for any time period in this study was based upon 31 observed utterances.

2.3 Dependent Variables

2.3.1 Conflict and co-operation by the Zimbabwe Government

To measure the changing strategies used by the Zimbabwe Government in its dealings with internal political opposition (including but not limited to the MDC), two separate variables – conflict and co-operation – were developed. Each variable measured actions taken by the government towards its opposition in the form of counts (e.g., the frequency of co-operative actions taken) across the initial 2007-2008 period of interest. This was achieved through the collection and coding of government actions from the website www.zimbabwesituation.com – a website that integrates all available news articles from online sources relating specifically to the situation in Zimbabwe¹⁷.

Actions taken by the Zimbabwe Government were identified from news articles across the span of the two-year period based on six criteria. These criteria were designed to identify actions that could be traced to a specific time and place¹⁸, and that were relevant to what was of interest –

¹⁷ Though independent data (from a non-governmental organisation for example) would have been preferred, the only data that could be obtained were for violent acts (collected by the Zimbabwe Peace Project) and these had already been aggregated into monthly frequencies. No data were freely available for the frequency of violence or co-operation at aggregates of less than a month.

¹⁸ Actions of the same type were considered to be separate if they were reported to have occurred in different places or on different dates. Conversely, all violence occurring in a specific incident happening in a single place

political government action with respect to political opposition. Thus, for a government action to be identified in a news article and be considered codable, it had to:

- (i) Have already happened. It must not have been an intended or planned event.
- (ii) Be traceable to a specific date *or* have occurred on every single day between two specified dates. (e.g., torture during a specified period of detention).
- (iii) Be a *concrete* action taken by the Zimbabwe Government and not rhetorical speech, promises, threats or speculation made by ZANU-PF party ministers.
- (iv) Be a purely politically-motivated¹⁹ action (not a result of the genuine running of the country).
- (v) Have been committed by the Zimbabwe Government *or* a member of any one of its surrogates. (Surrogates included ZANU-PF Government ministers, ZANU-PF party members, members of the police force, members of the armed forces, the group known as ‘War Veterans’²⁰ and members of the Central Intelligence Organisation)

could only be scored as a single event. Whilst this rule may not make intuitive sense in all circumstances such as in large battles occurring between countries, it appeared to work effectively within the confines of this internal crisis.

¹⁹ This was defined by how it was reported in the media (e.g., if the media reported stated or implied that the purpose of the action was political, such as an action aimed as disrupting an opposition political organisation, it was scored).

²⁰ The War Veterans are a group sponsored by the Government to perform rallies, intimidate political opponents and assist in takeover of commercial farms. Most are not actually veterans of the Rhodesian Civil war during the 1970s as is implied by the group’s name, with most having been too young to have served during this period.

- (vi) Have implications for those classified as internal political opposition to the Zimbabwe government (e.g., MDC members, political protestors of any kind, suspected MDC supporters, white farmers, critics of the government, and supporters of independent Presidential candidate Simba Makoni).

These actions were coded separately into action types using a coding system derived from the list of political actions described in the World Events Interaction Survey (WEIS; see Goldstein, 1992; McClelland, 1983). The WEIS list is a compilation of 61 types of action that can occur between political actors (e.g., states), and so provided a useful basis for generating a list of events that were applicable to the Zimbabwe crisis. After a number of discussions and test trials with coders familiar with events in Zimbabwe, the number of relevant action types was reduced to 15. These 15 were then grouped into five distinct action types as follows–

- (i) violent/aggressive actions against people
 - murders, injury/abduction, intimidation, arrest
- (ii) destruction/theft of resources/property
 - seizure of property/possessions, destruction of property, cutting off of aid/resources
- (iii) non-violent/non-destructive adversarial action
 - legal or formal complaints, cancelling or postponing co-operative events
- (iv) relief/cessation from/of adversarial action
 - returning property possessions, cessation of violence
- (v) pro-active co-operation
 - meeting to discuss issues or to negotiate, giving assistance, making substantive agreements

Three action coders trained together by going through a number of examples with the author, using this taxonomy to classify the actions by the Zimbabwe government collected from the news articles. They were instructed that actions could be coded for more than one type, but could only be coded for a single type once. For example, if war veterans killed one MDC supporter and assaulted another, whilst burning a property down (all in the same incident), this would be coded as two separate types of action (types 1 and 2) but only once for type 1, despite the fact that in this case two opposition supporters were affected by two different sub-types of violent/imposing action.

Given that the primary interest in the study was to develop a model predicting political aggression or co-operation, just two variables, the reader may wonder why five categories of action were used. The purpose of action-coders using this taxonomy was to enable them to differentiate actions of interest (i.e., acts of unambiguous aggression and co-operation – action types (i) and (v)) from those that could be classified as ‘adversarial’ but which were not especially aggressive or cooperative (e.g., actions such as legally banning journalists or releasing tortured opposition members). Hence, the primary function of scoring categories (ii) to (iv) was to filter out these actions and leave only the behaviour of interest.

2.4 Scoring Procedures

2.4.1 Conflict and co-operation events

The dates used for the identification of political actions were selected by random sampling throughout the 2007/2008 period until a reasonable proportion (judged by a visual inspection) of the years had been covered (approximately 18% of the days in each year). The person responsible for identifying actions taken by the Zimbabwe Government (in this case, the author) scanned all the news featured on the www.zimbabwesituation.com website for those selected

days²¹. Where gaps of 14 or more days occurred in the timeline without a sample date (after random sampling), the news for the median day within that time gap was also sampled. This occurred on four occasions in each year.

All actions by the Zimbabwe government that were identified as fulfilling the criteria for 'scorable' actions laid out above were copied and pasted into a Microsoft Word© document along with the date that the news was taken from and the date on which the action had occurred. This enabled the researcher to place the action onto a timeline after it had been coded.

Actions from random dates over the two year period were given to coders so that they could classify these actions into the five types outlined above. The agreement between the three independent coders on the classification of violent/co-operative events was substantial, with an agreement rate of 95%, for example, in classifying events as violent/imposing in a random sample of 108 events. After coding, all actions were put into a timeline developed in a Microsoft Excel© spreadsheet, where duplicated actions (i.e., the same actions reported and then later coded again²²) could be removed. Finally, all actions were summed using the different aggregations outlined below, with co-operative and violent actions added separately. The validity of the violence/imposition measure seemed to be more than acceptable as 2008 violence correlated strongly with *monthly*-aggregated violence data available for 2008 from the non-governmental organisation, the Zimbabwe Peace Project²³ ($r = 0.88$). However, no similar criterion for the validation of the co-operation measure could be found.

²¹ The number of articles for a single day varied significantly. This depended on the amount of reported news by media and not on whether the website had been properly updated. The website is updated daily and has been since 2000. The website can be viewed at www.zimbabwesituation.com.

²² This often occurred because news stories are constantly recycled across the year.

²³ See <http://www.kubatana.net/html/sectors/zim028.asp?like=Z&details=Tel&orgcode=zim028>

2.4.2 Psychological indicators – Computer scored

All the available material written by Manheru was scored for conceptual complexity, need for power and belief in ability to control events. This process did not require randomisation procedures, given that the computer software is unaffected by knowledge of the events occurring in the country or the hypotheses being tested.

2.4.3 Psychological variables – Hand scored

The two IC and power imagery scorers in this study were trained using the integrative complexity manual developed by Baker-Brown et al. (1992) and training materials provided by David Winter (see also Winter, 1991c). All scorers had to pass a standard IC or motivation imagery test where a minimum correlation of 0.85 with experts, as well a high level of agreement, had to be obtained. All scorers received material split into paragraphs and placed into a randomised order with indicators of time from the passages removed. For example, if the month of 'March' was mentioned in the scored material, this was replaced with [MONTH], indicating that Manheru had referred to a month and that this had been removed from the passage. Given that the changes in psychological indicators over time were of interest here, and not the differences between individuals, no action was taken to obscure the identity of the author of the passages or the people to whom he was referring.

For each variable, the primary scorer scored all the randomly selected material, whilst a secondary scorer scored approximately 10-20% of all the material scored by the primary scorer. Initial reliability checking (using the weighted version of the kappa statistic) suggested that the agreement was moderate to strong for both variables ($\kappa = 0.64$, for power imagery scoring and $\kappa = 0.57$ for IC, respectively), though this was not as substantial as previous studies.

2.4.4 Time aggregation

Finally, all correlational analyses were conducted using three different aggregations of the data – bi-weekly, tri-weekly and four-weekly periods starting at (and working back from) 1 January 2008.

This was done for three reasons. Firstly, as explained in the introduction, it has not yet been determined what the appropriate aggregation of time is when dealing with psychological indicators, even though intuition would dictate that we should look at shorter aggregates of time (i.e., given that psychological variables are more proximal indicators). Secondly, methodological work by researchers such as Shellman (2004) has shown that aggregating time series data into different intervals can dramatically shape one's interpretation of those data, so one should be careful not to make conclusions based on only one aggregation as though this applied to all aggregations. Examining the data aggregated into different time periods could therefore only enrich an understanding and interpretation of the data in this study.

Thirdly, being forced into choosing only a single data aggregation gives rise to a dilemma where, on the one hand, greater aggregation provides more accurately estimated data points, whilst on the other it gives rise to a loss of temporal resolution. At larger aggregates, more data are being used to estimate each value of the variables, which in turn means less measurement error. However, there may also be a loss of temporal precision with this increased aggregation. Genuine differences between two different time points at an aggregation of two weeks are no longer detectable when the data are aggregated into four week periods. The only way that this trade-off can be made optimal is to examine the data aggregated at different intervals.

3 RESULTS

3.1 Initial Data Analysis

There were five specific questions regarding the predictive use of psychological indicators in political crises that this work sought to answer. However, before describing the results that are directly relevant to answering these questions, it is useful to examine the effect that data aggregation had on the means and standard deviations of the variables used in this study. Summary statistics for each of the variables measured are displayed in Table 3.1.

This table helps to highlight the effect of using an increasing aggregation of time on the data analysis. For example, whilst the means of the psychological variables remain stable across the different aggregations and are thus relatively unaffected, the standard deviations become slightly smaller. In contrast to the psychological variables, the dependent variables show a rather different pattern. Here the means increase with aggregation (due to the same total number of events falling within fewer category boundaries) and standard deviations also increase. The increase in the standard deviations is due to the fact that in this dataset, violence and co-operation were measured as counts and the frequencies of these counts were unevenly distributed across time. Thus when creating a dataset out of larger units of time (i.e., four weeks instead of two weeks) many more actions went into some time periods than in others, making the difference between periods more extreme. Lastly, it should be noted from this table that the mean number and variability of violent acts committed by the Zimbabwe Government is far greater than the mean number variability of cooperative acts. Issues relating to this difference are treated in the discussion.

The first question of interest was whether the psychological indicators were related to aggression and/or co-operation in manner that has been shown in previous research. To examine this

question, multiple correlation analyses were performed with the two dependent variables and all six of the psychological variables. Before commencing with these initial analyses, the counts of co-operation and violence were root transformed to make any potential relationships between the psychological indicators and violence/cooperation variables more linear and also to help reduce heteroscedasticity (Cohen, Cohen, West & Aiken, 2003). This is not the optimal transformation for poisson-distributed count data (logarithmic transformations are), but given that at some aggregates there were a number of zero values²⁴, there were no other satisfactory options available at this exploratory stage of the analysis.

Table 3.2 shows the correlations between violence/co-operation with the psychological indicators occurring in the same time period at three different aggregations of the data. Looking across the three different aggregates, it is immediately clear that a number of variables were related to violence in the anticipated direction. This was particularly so at a data aggregation of three weeks, where four of the psychological indicators measured were significantly related to violence occurring in the same period of time. Greater (root transformed) violence in a time period was associated with greater power motive imagery ($r = .36, p < .05$), lower integrative complexity ($r = -.51, p < .01$), a more negative assessment of the political world ($r = -.37, p < .05$) and lower conceptual complexity ($r = -.39, p < .05$). All these relationships were in the anticipated direction. Meanwhile, the only significant relationship between (root-transformed) co-operation and a psychological indicator from the same period was with belief in ability to control events, which was significantly negatively related to co-operation ($r = -.39, p < .05$) when the data were aggregated into three week periods.

²⁴ The log transformation of zero results in an undefined number (negative infinity) and thus cannot be allowed in the analysis. Whilst some researchers recommend adding a small constant to the count data to avoid this problem (Mosteller & Tukey, 1977), there is no formal method by which to determine what this constant should be. Unfortunately, that decision is not a trivial one and as such adds another unnecessary element of arbitrariness to the data analytical process.

Lastly, with regard to the relationships exhibited between the computer-scored and hand-scored versions of power motivation and complexity, there seemed to be some evidence that the two complexity measures were related ($r = .45, p < .01$, when the data are aggregated at two weeks) but little evidence that the two power measures were measuring the same construct ($r = .20$, n.s., aggregated at two-week intervals).

In order to answer the second question of whether psychological indicators could *predict* violence and co-operation, additional correlational analyses were run, this time with all psychological variables lagged by one unit of time, so that values of psychological indicators were correlated with violence and cooperation occurring in the following time periods. The results of these analyses are shown in Table 3.3.

Predictive relationships with violence appeared for more than one aggregation of the data. The tri-weekly aggregated data showed that four variables were predictive of violence at a lag of one. These were power motive imagery ($r = .48, p < .01$), integrative complexity ($r = -.58, p < .01$) belief in ability to control events ($r = .39, p < .05$) and need for power ($r = .34, p < .05$) respectively.

At a data aggregation of four weeks the same significant relationships emerged, albeit more strongly. Integrative complexity was the best predictor ($r = -.76, p < .01$), followed by power motive imagery ($r = .60, p < .01$). Both variables were strongly predictive of violence. Belief in ability to control events and need for power were moderately predictive of the violence ($r = .46, p < .05$; $r = .45, p < .05$, respectively).

Lastly, it is worth noting that P1 and conceptual complexity were not found to be predictive of violence, even though they had been found to be related to violence during the same time period, whilst only need for power was predictive of co-operation ($r = -.32, p < .05$).

3.2 Fitting a Model to the Data

Given these significant results, the issue arose as to how a model should be fitted to the 2007 and 2008 data, in order to make real-time predictions in 2009. There were several choices available between four predictor variables and two different aggregations of the data. Because the predictive utility of psychological variables seemed to be strongest when the data were aggregated at four weeks intervals, a four week aggregation for the model was chosen. However, a more difficult choice arose with regard to which variables should be entered into the model, given that there were four apparently good predictors, two of which, need for power and power imagery, were supposed to measure the same underlying construct.

Interestingly, although need for power and power motive imagery were intended to measure the same dimension, they did not appear to be significantly related to each other at any data aggregation. Yet, to add to this confusion, both variables were related to violence. Conceptually at least, it would have made little sense to have both of these variables in the model and, given that power motive imagery appeared to be more strongly related to violence, it made more sense to drop need for power than power motive imagery. However, this was by no means a satisfactory end to the issue and it will demand further attention in the discussion.

Before entering the other three variables of interest into a regression model, the relationships between all three predictors and violence were examined visually. Given that, at the four week aggregation of the data, there were no zeros present in the data for the violence variable, violence was log transformed and then examined in relation to the three lagged psychological indicators.

Scatterplots of these relationships are shown in Figures 3.1 to 3.3. All have a non-parametric curve fitted and all three appeared to have sufficiently linear relationships with the log transformed version of violence.

A simultaneous entry ordinary least squares²⁵ regression model was therefore tested with belief in ability to control events, integrative complexity and power motive imagery (all at lag of one) entered as predictors of log-transformed violence. The overall model was highly significant, $F(3, 19) = 17.49, p < 0.001$, accounting for 73% of the variance in log-transformed violence ($\text{adj}R^2 = .73$). Moreover, two of the predictors, integrative complexity and power motive imagery, both provided unique predictive utility in the model, $\beta = -.55, t(19) = -3.74, p < 0.002$, and $\beta = .32, t(19) = 2.41, p < 0.03$, respectively. Belief in ability to control events did not reach a statistical level of significance, $\beta = .23, t(19) = 1.71, p = 0.11$, but converged towards significance²⁶. These figures are repeated for the reader's convenience in Table 3.4.

Given that the dependent variable in this regression analysis was derived from transformed count data that were also serial (i.e., time-series) in nature, heteroscedasticity and autocorrelation could both have been a concern with respect to making the correct statistical inferences. However, a Breusch and Pagan²⁷ Test for heteroscedasticity, $X^2(4, N = 23) = 1.64, p = .20$, and a Durbin-Watson test of autocorrelation ($d = 1.50, p = 0.16$) revealed that there was no significant evidence of either problem. It could therefore be reasonably concluded that power motive

²⁵ Generalized Linear Models are sometimes a preferred option for testing models with Poisson-distributed dependent variables. However, in this case an ordinary least squares regression was preferred because of the well-developed convention of calculating R^2 . Though conventions do exist for calculating variance explained in generalized linear models, there is no fully accepted convention. Since the ordinary least squares regression model that was tested did not exhibit non-normality of residuals or significant heteroscedasticity, there was no reason for alarm in opting for a more traditional approach.

²⁶ The p values shown are for two-tailed tests. Though one might justify one-tailed tests, given the directional predictions, the number of choices available in fitting the model meant that a statistically more stringent criterion was more appropriate.

²⁷ The Breusch and Pagan (1979) test was also backed up by a visual inspection of fitted values against the residuals and also a visual inspection of the normality of the residuals. Neither inspection revealed any obvious issues.

imagery and integrative complexity together provided significant and strong predictive utility in predicting violence when aggregated over four-week periods.

3.3 Out of Sample Forecasting

As stated in the introduction, the intention was to use an emergent model from the historical data for real-time forecasts in 2009. This is an extremely important part of testing the forecasting properties of a model and without it the utility of a model remains uncertain. In this study, accomplishing real-time forecasts would have also spoken to the fifth question specified in the introduction of whether or not this predictive psychological model of adversarial intent could provide accurate predictions for current crisis behaviour. Unfortunately, due to unforeseen circumstances, this proved impossible. In February of 2009, the columns of Nathaniel Manheru stopped appearing in *The Herald* newspaper.

Faced with situations where real-time forecasting is impractical, some researchers have used a pseudo out-of-sample forecast, where the sample is split with the first part of the sample used to fit a model for predicting the second part (Chatfield, 2001). Unfortunately, the study's sample size was far too small to do this, especially at the four week aggregation for which the model was eventually developed ($n = 11$ or 12 , when split in half). Additionally, one could reasonably take the view that this method is too open to manipulation (for instance, by splitting the sample at a favourable point) to be a truly useful way of testing the forecasting model as evidence for others to fairly evaluate.

For these reasons, these particular results were unable to provide an answer to the pivotal question of how the model fitted in this study would perform in real-time. However, to provide at least some evaluation of the predictive utility of the model, a time-series of violence against predicted violence (i.e., violence predicted by the fitted model) is provided in Figure 3.4. As is

apparent from this figure (if it were not already so from the regression analysis) the model is an exceptional fit for the data. Unfortunately, this does not represent a true forecasting test and should not be treated as such. But it is useful to note, just from this time series, that the predictions are not perfect. At the beginning of the time series, the model fails completely to predict the surge in violence in Period 3. This issue is dealt with extensively in the discussion.

Table 3.1. Summary statistics for all variables examined at each aggregation of time

	Two Week Aggregate			Three Week Aggregate			Four Week Aggregate		
	mean score	standard deviation	sample size	mean score	standard deviation	sample size	mean score	standard deviation	sample size
Power Imagery	16.36	3.26	40	15.70	3.03	30	16.24	2.86	25
P1	0.14	0.16	40	0.13	0.16	30	0.12	0.14	25
IC	1.86	0.31	40	1.86	0.30	30	1.86	0.28	25
BACE	0.27	0.07	35	0.27	0.07	29	0.27	0.07	24
CC	0.62	0.04	40	0.62	0.04	30	0.62	0.04	25
nPower	0.19	0.09	35	0.17	0.08	28	0.18	0.08	24
Violent Acts	7.79	9.47	52	11.88	12.42	34	15.58	16.09	26
Co-operative Acts	0.73	1.30	52	1.12	1.68	34	1.46	2.08	26

P1 - Belief about the co-operativeness of the political world, IC - Integrative Complexity, BACE - Belief in Ability to Control Events, CC - Conceptual Complexity, nPower - Need for power

Table 3.2. Inter-correlations between variables at different aggregates of time

Variable	1	2	3	4	5	6	7	8
Aggregated at two week intervals (35 ≤ n ≤ 52)								
1. PMI	-	-0.38**	-.37**	-.036	.00	.20	.19	.10
2. P1		-	.08	.02	-.04	.19	-.34*	-.11
3. IC			-	-.11	.45**	-.07	-.47**	0.23
4. BACE				-	.10	.26	.06	-.12
5. CC					-	.24	-.17	.09
6. nPow						-	-.04	-.00
7. vViol							-	-.08
8. vCo-op								-

Aggregated at three week intervals (28 ≤ n ≤ 34)								
1. PMI	-	-.58**	-.35*	.02	-.24	.19	.36*	-.03
2. P1		-	.10	.14	.15	.22	-.37*	-.01
3. IC			-	-.39*	.30	-.11	-.51**	.25
4. BACE				-	-.13	.17	.21	-.31
5. CC					-	.08	-.39*	.00
6. nPow						-	-.03	.12
7. vViol							-	-.17
8. vCo-op								-

Aggregated at four week intervals (24 ≤ n ≤ 26)								
1. PMI	-	-.49**	-.45*	.03	-.16	.17	.23	-.12
2. P1		-	.06	.13	.08	.40*	-.16	.17
3. IC			-	-.40*	.32	-.31	-.62**	.18
4. BACE				-	-.17	.25	.33	-.39*
5. CC					-	.12	-.33	-.14
6. nPow						-	-.03	-.12
7. vViol							-	-.13
8. vCo-op								-

Note. * - p<0.05, ** - p<0.01, one-tailed tests. PMI - power motive imagery, P1 - beliefs about the cooperativeness of the political world, IC - integrative complexity, BACE - belief in ability to control events, CC - conceptual complexity, nPow - need for power, vViol - square root of violence, vCo-op - square root of co-operation.

Table 3.3. Inter-correlations between variables at different aggregates, with psychological variables at lag 1.

Variable	1	2	3	4	5	6	7	8
Aggregated at two week intervals ($34 \leq n \leq 52$)								
1. PMI	-	-.38**	-.36*	-.03	.01	.20	.25	-.17
2. P1		-	.09	.03	-.03	.19	.05	.23
3. IC			-	-.14	.44**	-.06	-.45**	.19
4. BACE				-	.08	.28	.24	-.21
5. CC					-	.26	-.29	-.17
6. nPow						-	.09	-.05
7. vViol							-	-.08
8. vCo-op								-
Aggregated at three week intervals ($27 \leq n \leq 34$)								
1. PMI	-	-.57**	-.33*	.03	-.24	.18	.48**	-.05
2. P1		-	.06	.12	.15	.23	-.13	-.09
3. IC			-	-.42*	.30	-.11	-.58**	.05
4. BACE				-	-.13	.17	.39*	-.27
5. CC					-	.08	-.16	-.14
6. nPow						-	.34*	-.32*
7. vViol							-	-.17
8. vCo-op								-
Aggregated at four week intervals ($23 \leq n \leq 26$)								
1. PMI	-	-.48**	-.44*	.05	-.14	.16	.60**	-.19
2. P1		-	.04	.12	.07	.42*	-.16	.04
3. IC			-	-.44*	.30	-.30	-.76**	-.08
4. BACE				-	-.18	.26	.46*	-.11
5. CC					-	.13	-.19	-.07
6. nPow						-	.45*	-.02
7. vViol							-	-.13
8. vCo-op								-

Note. * - $p < 0.05$, ** - $p < 0.01$, one-tailed. PMI - power motive imagery, P1 - beliefs about the cooperativeness of the political world, IC - integrative complexity, BACE - belief in ability to control events, CC - conceptual complexity, nPow - need for power, vViol - square root of violence, vCo-op - square root of co-operation.

Table 3.4.
Summary of Regression Analysis for Variables
Predicting Log Violence at a Lag of 1 (N=23)

Variable	<i>B</i>	SE <i>B</i>	<i>β</i>
Power motive imagery	0.11	0.04	0.32*
Integrative complexity	-1.89	0.51	-0.55**
Belief in ability to control events	3.2	1.87	0.23

Note. $\text{adj}R^2 = 0.73$, ** - $p < .01$, * - $p < .05$

Figure 3.1 Power Motive Imagery plotted against Log-transformed Violence. A locally weighted scatterplot smoother line is fitted.

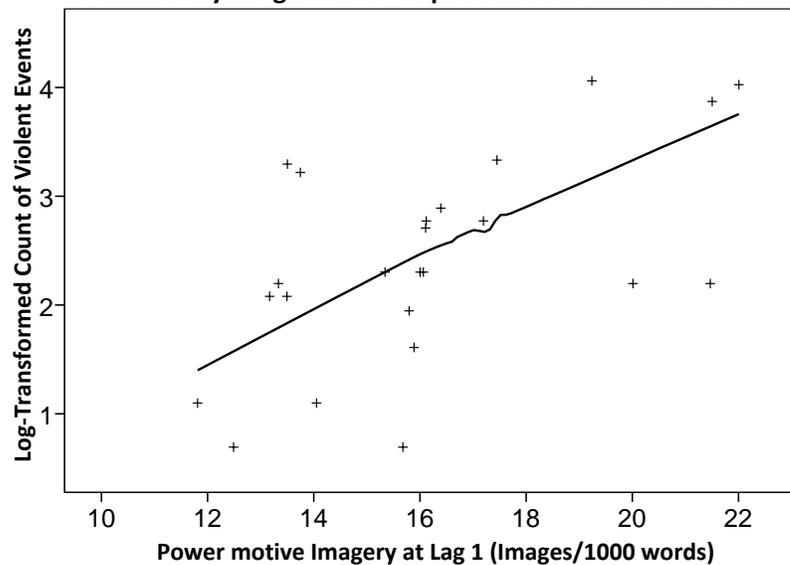


Figure 3.2. Mean Integrative Complexity plotted against Log-Transformed Violence. A locally-weighted scatterplot smoother line is fitted.

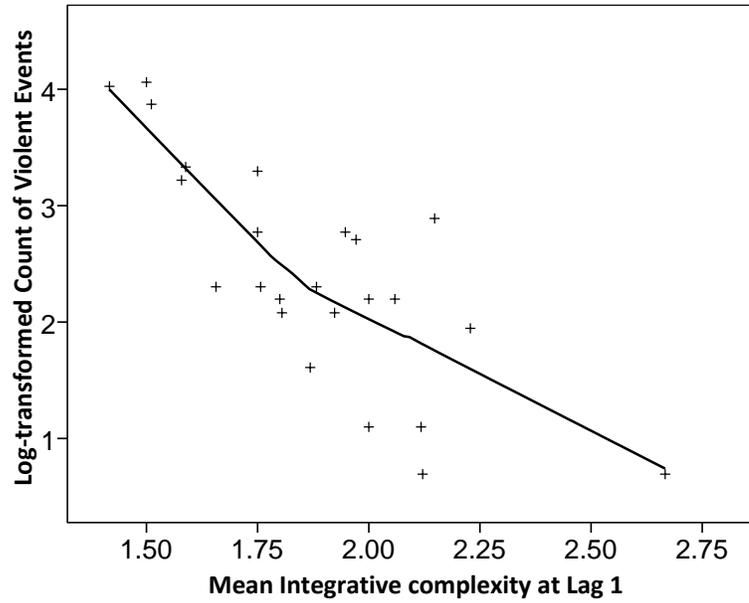


Figure 3.3. Log-transformed Violence plotted against Belief in Ability to Control Events. A locally weighted scatterplot smoother line is fitted.

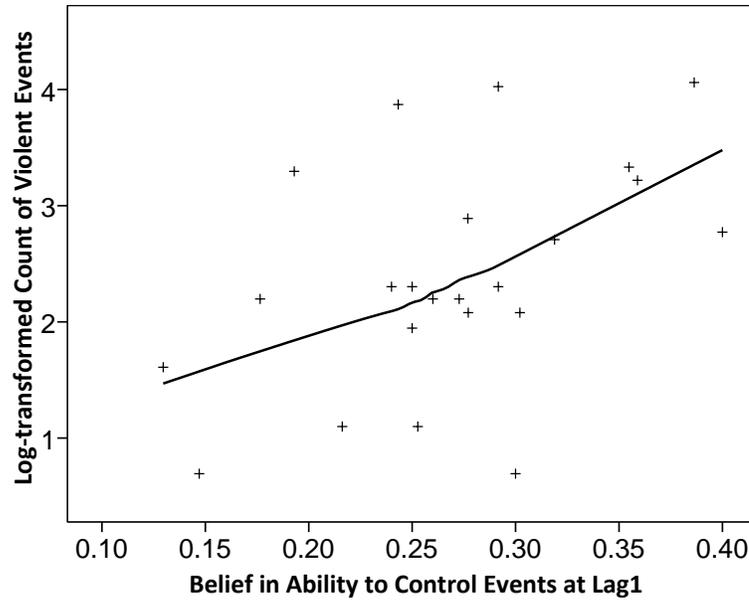
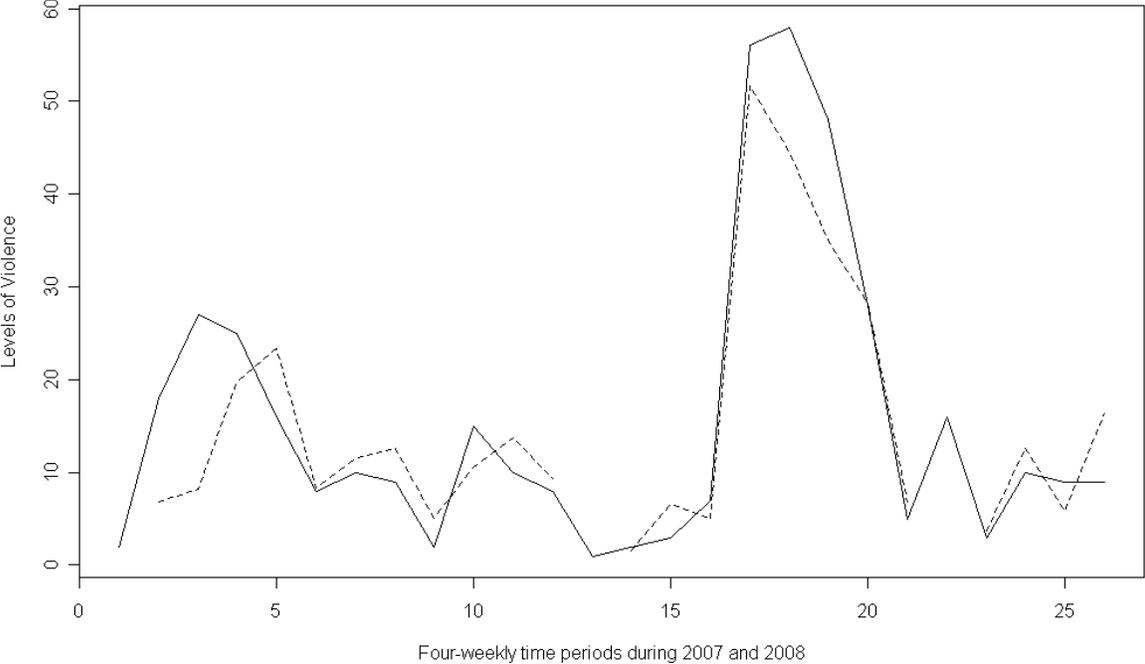


Figure 3.4. Time series of Violence and Predicted Violence by the Fitted Model. Dotted lines are predictions.



4 DISCUSSION

4.1 Evaluation of the Central Findings

This research sought to answer several questions regarding the predictive utility of tracking psychological indicators during a political crisis. It attempted to do this by retrospectively tracking the crisis situation in Zimbabwe for the years of 2007 and 2008 and using a stochastic model derived from this historical period to accomplish real-time forecasts. Fitting a time-series regression model to the historical data helped to provide answers to four out of the five questions posed in the introduction. However, it was not possible to answer the fifth question because using the model for real-time forecasting was not possible in this case. The answers to these questions that could be provided now form the basis for the following discussion.

4.1.1 Question 1

Are psychological indicators of intent related to aggressive/cooperative acts by the Zimbabwe Government? The answer to this question with regard to the relationship between the psychological indicators and aggression was a conclusive 'yes'. All of Manheru's psychological indicators showed some significant relationship with government violence for at least one of the aggregate/lag combinations. The most consistent of these was Manheru's integrative complexity, which showed a moderate to strong relationship with government violence when examining all aspects of the dataset. Unsurprisingly, this was closely followed by Manheru's power motive imagery, which also appeared to be substantially related to government violence across a number of aggregations and at both lags of the variable. This meant that the two psychological indicators that have most often been linked to political aggression in the past were shown again to be the best indicators of aggression during the Zimbabwe crisis.

The results for the other variables were more mixed. Manheru's conceptual complexity, for instance, appeared to be related to violence only at a lag of zero and specifically at the three week aggregation, whilst the performance of other variables could be placed somewhere between that of conceptual complexity and power motivation imagery. All of these relationships should be treated with some caution, given the large number of correlations being examined (i.e., one must always expect some significant relationships by chance alone when performing so many analyses). However, it was encouraging to have at least partially replicated all the findings from previous studies and especially encouraging that none of the relationships that were found ran counter to the direction predicted.

In contrast to the relationship between psychological indicators with aggression, the relationship between psychological indicators and cooperation appeared to be non-existent. Only two significant correlations were found between psychological indicators and cooperation out of a possible forty-two – precisely what one would have expected by chance alone.

There are a few possibilities as to why psychological indicators did not have the expected relationship with co-operation when they did exhibit the expected relationship with violence. The first is that there simply wasn't enough cooperation over the 2007 and 2008 period for the psychological indicators to detect – a restriction of range problem. This is supported by the fact that the mean and standard deviation of the cooperation variable were far smaller than the mean and standard deviation of violent acts over the same period. A second possibility is that the measure of co-operation used may not have been sensitive enough to the cooperation that was occurring. During periods in 2007 for example, there were sometimes long periods of media blackout while talks between ZANU-PF and MDC representatives took place, meaning that not enough information would have been filtering through to news reports to be identified by the action identification procedure (see Method section). Lastly, it is possible that none of the so-

called acts of co-operation measured were genuine, at least in the mind of Nathaniel Manheru, meaning that one did not see the associated psychological changes. This view is supported by the fact that cooperation and violence were only slightly negatively correlated, signifying that ZANU-PF violence was still ongoing during co-operative periods – surely a sign of bad will on the part of ZANU-PF. It is reasonable to suggest therefore that the talks and agreements that took place could have been a façade as far as the intentions of ZANU-PF were concerned, thus making it very difficult to fit a predictive model. After all, if psychological indicators are supposed to reveal the *true* intent of a leader or government, one would only expect them to change in relation to genuine acts of cooperation.

Whilst this latter argument might seem circular (i.e., ‘if the psychological indicators don’t work, then one can just put it down to the dependent variable being incorrect’), it should be noted that cooperation can be easily feigned compared to violence, which either occurs or does not occur. Had the same lack of relationship occurred between violence and the psychological indicators, it would have been difficult to justify this same line of argument. In future therefore, it may be necessary to create measures of cooperation that are resistant to ambiguity – perhaps by limiting the definition of co-operation only to instances where substantial compromise or agreements are made.

4.1.2 Question 2

In addition to being related to violence or co-operation, can these psychological indicators predict violent and co-operative acts? This question, with regard to co-operation, is now redundant given that it was found that co-operation was not consistently related to any of the psychological indicators. However, with regard to the prediction of aggressive government actions, a number of patterns emerged from the data that are worth addressing.

Firstly, it was clear that some variables were related to violence in the same period that the violence was occurring but were not predictive of violence. For example, belief about the cooperativeness of the political world (P1) was related to violence at aggregates of two and three weeks, but this relationship disappeared completely when violence was correlated with P1 occurring in the period before. Conceptual complexity also showed a similar pattern. In the immediate context of this study therefore, it seemed that these variables did not serve as a useful part of predictive model, though this does not rule out the possibility that these variables are still predictive of violence - it could be that the aggregations used in this study did not offer enough temporal precision for any predictive utility to be detected.

Most of the variables, however, were more predictive of violence than they were indicative. Integrative complexity and power motive imagery were related to violence at both a lag of zero and a lag of one, but these relationships were more substantial at a lag of one. Need for power and belief in ability to control events were not related to violence occurring in the same period but were predictive at aggregates of both three and four weeks.

In sum, there was clear evidence showing that these variables could be used in a predictive fashion. Though this has been shown previously to some degree, in work by Suedfeld and Bluck (1988, integrative complexity) and Winter (1993, power motive imagery) for example, no previous published work has addressed the predictive utility of belief in ability to control events or need for power within the tracking approach framework - both are normally viewed as traits and used in profiling work (e.g., Hermann, 1980; Hermann, 2006; Winter et al., 1991). In this respect, these findings begin to fill a gap in the literature regarding the use of these trait variables to detect changes over time and make predictions about behaviour based on those changes.

4.1.3 Questions 3 and 4

Over what time-frame are psychological predictions of violence possible and/or optimal? How much variance in political strategy can be predicted from a tracking approach to prediction (i.e., just how useful is the predictive utility)?

Questions three and four from the introduction are best answered together and in the order just written. The time-series regression model that was ultimately tested was derived from the sample using the four week aggregation, as it was at this aggregation of the data that the psychological indicators seemed to be most predictive of violence. Within this particular dataset therefore, the answer to this question was that the predictions made at a four week aggregation of the data were optimal.

However, this may not always be so. The precision with which one estimates the psychological indicators and government actions improves greatly when one increases data aggregation. This is the most likely reason that the four-weekly aggregation period proved to be the best in this instance and produced the strongest correlations. However, if greater precision can be achieved at shorter intervals (e.g., by scoring more material, by scoring more reliably or by gaining access to comprehensive high resolution conflict data), this may yet prove to be the better option, especially given that psychological changes can happen rapidly (e.g., as found in Chiapas negotiations, Liht, Suedfeld, & Krawczyk, 2005).

In answer to the next question, it seemed that the model fitted at the four week aggregation worked extremely well. Nearly three-quarters of the variance of log violence was predicted by the combination of power motive imagery, integrative complexity and belief in ability to control events, suggesting that these psychological factors together (but integrative complexity and power motive imagery in particular) could form the basis of the key psychological 'formula' to

predicting real-time events. The model is certainly the biggest success of this particular study and will no doubt be the subject of further investigation.

However, there is a good statistical reason why one might be a little sceptical that the model explained as much variance in violence as it apparently did. It should be noted that in the course of the model building there were a large number of choices that could have been made with different aggregations of data and a number of different variable combinations. This was made as transparent as possible by the manner in which the results section was handled (i.e., by not obscuring the many possibilities available). As with the correlation analyses, there is a risk of reading ‘too much’ into a model that was derived from a large number of initial possibilities and one should always expect some favourable results by chance. If the data had only been aggregated into two-week periods, a rather different set of conclusions might have been reached from this study.

At the same time, one can take comfort in the fact that a model could have also been fitted at the tri-weekly aggregation of the data. The correlations between predictors and violence at this aggregate were not trivial. In sum, it can probably be safely concluded that the psychological indicators in this study do provide a solid basis from which one could make real-time predictions.

4.1.4 Question 5

Does a model derived from the historical data extrapolate to making predictions made in real time, or alternatively, ‘out of sample’? Of the five questions posed in the introduction, it is unfortunate that it was the last one that could not be answered. Many of the uncertainties regarding the answer to previous questions (such as the variance in violence predictable from psychological indicators) could be resolved with an extended period of out-of-sample forecasting. Due to reasons already stated, an extended period of forecasting was impossible.

However, this failure is also an opportunity to raise an obvious but important question— could this problem have been avoided by, for example, scoring more individuals or scoring a combination of materials?

Although this is a perfectly reasonable question to ask, scoring more individuals or a combination of materials would have raised as many issues as it solved. Being able to track indicators over time and interpret those changes relies on holding a number of factors constant. A symposium on content analysis in the 2000 edition of the journal *Political Psychology* identified a number of factors that can adversely affect content analysis scores, including whether transcripts are from spontaneous or prepared statements (Dille, 2000), whether or not materials are from the same source (Dille & Young, 2000) and whether the sources are private or publicly available (Marfleet, 2000). If one scored material from a large number of different sources and different people, it would be difficult to keep track of all these different potential issues, and although this has been done in a number of studies retrospectively (i.e., where it didn't matter what material might be available in future), one could not control the relative availability of different types of material in real time (except perhaps when resources are abundant) and this could lead to systematic and disastrous forecasting biases, should this availability change over time.

Thus, with regard to Zimbabwe, scoring Manheru's articles was the only really viable option. Whilst one could have scored the snippets of information found on the internet on occasion, it would have been a contamination of the dataset to include these with the material from Manheru and unjustifiable given the wide availability at the time of Manheru's newspaper columns. It made sense then that Manheru should be scored alone, particularly given that the column had been running since well before George Charamba took over the reins 2005 and had showed no sign of abating. The departure of Manheru could not reasonably have been foreseen.

However, although an examination of the in-sample forecast featured in Figure 3.4 is certainly not as useful as an out-of-sample forecast in determining the utility of the model (forecasting models never perform as well out-of-sample, Chatfield, 2001), it is worth discussing its success in predicting the large wave of violence occurring in Period 17 in comparison to its failure to predict another surge in violence in Period 3 (refer to Figure 3.4). What was different about these two periods that the model was late in predicting the first surge and near perfect in predicting the second?

To answer this question, it is helpful to place these events in context. During March 2007, (which approximates to period 3 in Figure 1) the surge in violence occurred in the immediate context of a number of gatherings and protests by the MDC and other political organisations within the country that oppose ZANU-PF. The Zimbabwe Government, in reaction to this increase in the frequency of large gatherings reacted aggressively and *quickly*, arresting many leading MDC members and having them beaten in detention. In contrast, the second large wave of violence began after the March elections in 2008, when the MDC had just achieved a surprising victory in the parliamentary elections and a majority of the vote in the presidential election. This time, however, the violence was not an immediate consequence of these events. For weeks, the government held back the presidential results and only allowed parliamentary results to trickle through, whilst the country remained peaceful. Eventually however, following the announcement of the results and the declaration of the next election in June, vicious violence against opposition supporters began unabated.

It seemed that the failure to predict in the first instance may have been the immediacy of the response to the ‘provocation’ (placed in inverted commas because these gatherings were not at all aggressive) provided by the protests in 2007, compared to the rumination that seemed to take place after the March election in 2008. It could be that the model (aggregated at four weeks)

may only work for *planned* aggression, where intent takes some time to be put into action. This would certainly explain the failure to predict the first wave of violence and would also suggest a need to look again at the possibility of using shorter time intervals over which to measure psychological change.

Unfortunately, in-sample forecasts do not give researchers much more to work with than what has just been described. However, though they provide little in terms of evaluating the tracking model developed here, one can still hypothetically address potential performance issues that may have arisen, had the model been used during the course of 2009. For instance how would this tracking approach to prediction have performed in circumstances of changing norms within the country and, more specifically, what would happen to the predictive utility of the model if members of ZANU-PF suddenly felt restricted in the extent to which they could use violence to suppress political opposition?

This is an important question and, had forecasting continued into 2009, it would have become relevant by the end of February when the MDC and ZANU-PF parties formed a Government of National Unity (GNU). This agreement meant that Tsvangirai became Prime Minister of Zimbabwe whilst his second in command, Tendai Biti became Finance Minister, thereby bestowing upon them a new level of legitimacy that they had never before achieved. Given this change in the political landscape, it may have not been possible anymore (even if it had wanted to) for ZANU-PF to still openly beat opposition members. To do so would have been in defiance of the expectations laid out by the Southern African Development Community, to which ZANU-PF had become at least partly accountable.

This level of change in circumstances could in itself doom a tracking approach model *if* one were to make purely data-driven predictions in the new context. However, if used intelligently, the

model could still be utilised in a new context and be a useful guide. For example, following the creation of the GNU, one may not have predicted widespread violence in Zimbabwe anymore when the indicators suggested this, but one might have predicted more covert forms of aggression such as attempted assassination attempts (disguised as accidents for example) or unexplained disappearances of opposition members.

There have already been a number of suspicious car accidents involving MDC officials - one in which the wife of Morgan Tsvangirai was killed²⁸ - and there is no reason why the model could not provide an indication that something of this more subtle nature was about to occur, even if the predicted values from the model (i.e., acts of violence) might seem nonsensical when interpreted literally. One might instead predict that some major 'incident' of aggression was about to occur, one which may have great political impact (such as an assassination) but which would ultimately be scored as a single act of violence on the aggression scale²⁹. In sum, changing norms would not make tracking approaches completely redundant.

In addition to addressing the problem of changing norms, one might also ask the question of how a model such as this might fare in other crisis situations, ones in which there were many complex factors at work. It was fortunate, for instance, that the Zimbabwean situation was simplified by the dominance of ZANU-PF in the country and the comparative weakness of the MDC. Much of the political interaction in the country during 2007 and 2008 was unidirectional in that ZANU-PF had control of the police, the army and through them also the general population. Moreover, it was clear that ZANU-PF had made it their priority to stay in power, making its psychological

²⁸ This is still officially considered an accident. However, since the crash occurred, it has since become known that the USAid truck which collided with the Tsvangirais' vehicle, was actually being driven by a person not registered with the Aid organisation. The case is still under investigation.

²⁹ Of course, statistically validating the forecasts in these situations would be impossible, because even a successfully predicted high-level assassination would appear a failure given that the predicted number of violent incidents was large.

indicators (as measured through the Manheru column) most likely to change in relation to achieving that goal. This combination of factors made the psychological modelling of the situation there considerably easier.

In other situations however, interpreting changes in psychological indicators may be treacherous due to the complex set of dynamics taking place. For instance, in democratic countries, leaders not only face the toll of dealing with foreign policy crisis issues but also have to concern themselves with their accountability to the electorate. Winter (1996), for instance, has found that Bill Clinton evidenced large increases in his power motivation during the second two years of his first term in office. Though this was interpreted by Winter as a reaction to increasing domestic political pressure it may have also been related to, for example, the military action being taken in Haiti in 1994 or Yugoslavia in 1995 (Suedfeld, Cross & Stewart, 2008). Thus the occurrence of multiple independent events that could affect psychological indicators in a short time span is one factor that could make the interpretation of a tracking approach to prediction somewhat precarious.

In the long-term, the only way that these questions will be answered is through the careful testing of tracking models in real time. Unfortunately, this will not be immediately possible for Zimbabwe, but it may be possible for other crises in due course. In any case, and regardless of the fact that this study was unable to produce out-of-sample forecasts, the approach still remains promising and still offers many advantages over the profiling approach that has been dominant in political psychology so far. What is needed now is further validation.

4.2 Further Issues, Problems and Criticisms

So far, this discussion has been focussed on the central questions posed at the end of the introduction. But there are also other issues one might address that are not central but which are nevertheless important. These are now addressed, each in turn.

4.2.1 Weaker than normal reliability

One such issue is the reliability with which power motive imagery and integrative complexity were scored. Though the reliability was moderate to strong in absolute terms, experienced researchers in this area will know that these variables are usually scored with a substantial degree of reliability (usually $r = .80$ between coders). The likely reason for the attenuated reliability is in part due to regular human error in scoring large amounts of material but is also likely due to the unusual scoring material, which the secondary coders (who were less familiar with the context) commented on as being quite irregular.

Manheru's (or rather George Charamba's) first language is not English and his writing style is crude and disjointed. Many of his passages are rich in subtle innuendoes. Thus, there were many cases in which scores were assigned with some uncertainty and removing these ambiguous passages from the dataset may improve the reliability immensely (as is often recommended in dealing with ambiguous material) but at a cost in scorable material. Finding this balance is difficult. However, the presence of less than ideal reliability casts an unnecessary shadow over the validity of the findings here and might be improved upon. Doing so would not only reassure would-be critics of the external validity of the study, but improve results at lower aggregates of time by enhancing the estimation of variable values.

4.2.2 Practical considerations

Another issue is whether the tracking approach to prediction is feasible or economic on a practical level when used in real-time. For instance, there is a substantial amount of work involved in the tracking approach to prediction that is not required in the profiling approach practiced by Winter et al. (1991a) and Winter (2001; personal communication, April 24, 2009). This is because the tracking approach requires constant scoring whereas a profile, once established, does not require further information. With limited resources available, it stands to reason that a tracking approach used in real time might not be achievable based on the work of a small group of academic researchers, though this remains to be seen.

For this reason, the issue of whether the computer programs are as good at capturing psychological dimensions as human scorers is especially important, given the efficiency with which computers can score large amounts of material. Unfortunately, in this study, it seemed that computers were not the answer. On one hand, and very encouragingly, two of the computer-scored variables – belief in ability to control events and need for power – were both predictive of violence. On the other hand, both these variables were outperformed by the two hand-scored variables.

In addition, neither conceptual complexity nor need for power seemed to behave similarly to their hand-scored counterparts. This is somewhat more understandable in the case of conceptual complexity (which only attempts to measure differentiation and not integration) but need for power and power motive imagery have the same definitions³⁰.

One reason why this might be so is that the computer-scored variables have been designed to score only spontaneous (i.e., not written or prepared) materials. Hermann (2006) has always

³⁰ This is despite the fact that one is scored more as a trait (need for power) and the other is scored as a motive – a trait and a mental state.

insisted on scoring spontaneously uttered remarks by political leaders and it stands to reason that the programming may not work as effectively for the kind of language found in newspaper columns, *if* the program's decision rules have been written specifically for language use in that different context. It could also be the case that Herman's (1980, 2006) version of need for power may grasp at the same construct as power motive imagery from a different angle, but one would still expect a strong relationship between variables that are defined in precisely the same way. Thus need for power, though potentially a very useful variable for the tracking approach to prediction, may need to be re-defined in some way if it is to be distinguished properly from power motive imagery.

In sum, the computer-scored variables offer an opening more than a solution to problems of practicality. However, until more intelligent machines are built, the field may be waiting a longer period of time than it would like for the valid coding of hand-scored variables by computers and, consequently, the solution to problem of inefficient coding.

4.2.3 Other approaches to predicting conflict: Where a tracking approach fits in

Lastly, given the large amount of variance in violence explained by the model developed in this study, there may be some researchers who remain sceptical of a psychological model providing so much predictive power, especially given that other factors (e.g., infant mortality, military build-up, threats to security, economic factors, religion, political systems, identity etc) that are connected to political violence are not even featured in the model. How, for instance, could psychological factors possibly be better at predicting crisis behaviour than economic or national security concerns?³¹

³¹ This question seems to be implied when people have asked the author about including other factors (security threats, troop build-up) in the predictive model. To this it should be said that these other factors are already implicitly in the model in that the ultimate effects of these factors on the psychology of the decision-makers is being captured.

It is important to point out therefore that those who would model these extraneous factors as *competing* with psychological factors may in fact have misinterpreted the purpose of developing a tracking approach to prediction. The belief driving the investigation of the tracking approach is not that psychological factors are more important than none-psychological factors in terms of explaining aggressive crisis behaviour. Rather, the assumption is that for all those other important factors to exert their effect on crisis behaviour, they must first effect changes in the psychology of the decision-makers.

Psychological change is the most proximal predictor of a change in crisis behaviour and one should expect therefore that proximal indicators would be strongly related to behavioural outcomes - perhaps even more than these more distal factors. Indeed the data in this study support the view that psychological indicators can account for a large proportion of variance in crisis behaviour.

Conversely, one may have to go beyond a psychological *tracking* approach if one is searching for longer-term predictors of conflict (e.g., profiling approaches). The psychological changes that have been measured here are proximal and only proximal³², meaning that one's predictions in a model of this kind are only as good as the last measurement made. It is therefore in the longer-term context that research into environmental change, societal change and political systems is ideally placed and is unlikely to be supplanted by a psychological tracking approach to prediction.

³² Winter's (1993) study on Annual Sovereign speeches is clearly an exception here, but an exception it remains. Most studies interested in psychological change before behaviour occurs have sought (and found) meaningful changes over shorter time periods (e.g., Suedfeld & Bluck, 1988) rather than over a year or more.

4.3 Conclusion - The Tracking Approach to Predicting Adversarial Intent

This study has shown that tracking changes in multiple psychological factors over time could provide a powerful way of predicting crisis behaviour. It has not only replicated previous findings that link various psychological indicators to political aggression, but it has also been able, for the first time, to combine these into a model with the explicit purpose of predicting aggression in real time. Moreover, this model, based chiefly on the contribution of two predictors (integrative complexity and power motive imagery) appears to account for large portions of the variance in crisis behaviour – more than adequate to provide sound estimates of future crisis behaviour.

However, there are also reasons to temper the claims that can be made directly from this study alone. Question marks remain over how such a model will extrapolate with changing norms or be useful in more complex situations, whilst weaker than normal reliability and practical considerations weaken any claim that a tracking approach to prediction (based on hand-scored thematic content analysis only) provides the ultimate solution to making short-term predictions regarding crisis behaviour. Even more importantly, the study was unable to produce out-of-sample forecasts that would provide a crucial test of the approach. Until this occurs, the exact potential of the tracking approach to prediction will remain unknown and unfulfilled.

Lastly, a model of this kind, even if it does prove to be an excellent predictive tool, is unlikely to replace other approaches to predicting conflict that emphasise long-term causal factors. The predictions made by the tracking approach are short-term in nature and this is most likely where its potential utility lies, though further research will certainly be required to explore just what that utility is.

REFERENCES

- Axelrod, R. (1973). Schema theory: An information processing model of perception and cognition. *American Political Science Review*, 67, 1248-1266.
- Axelrod, R. (Ed.) (1976). *Structure of decision: The cognitive maps of political elites*. Princeton: Princeton University Press.
- Baker-Brown, G., Ballard, E.J., Bluck, S., de Vries, B., Suedfeld, P. and Tetlock, P. E. (1992). The conceptual integrative complexity scoring manual. In C.P. Smith. (Ed.), *Motivation and personality* (pp. 400-418). Cambridge: Cambridge University Press.
- Breusch, T. S. and Pagan, A. R. (1979) A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47, 1287–1294.
- Chatfield, C. (2001). *Time-series forecasting*. Boca Raton: Chapman & Hall.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioural sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Crichlow, S. (1998). Idealism or pragmatism? An operational code analysis of Yitzhak Rabin and Shimon Peres. *Political Psychology*, 19, 683-706.
- Crichlow, S. (2005). Psychological influences on the policy choices of secretaries of state and foreign ministers. *Cooperation and Conflict*, 40, 179-205.
- Dille, B. (2000). The prepared and spontaneous remarks of presidents Reagan and Bush: A validity comparison for at-a-distance measurements. *Political Psychology*, 21, 573-585.

Dille, B. & Young, M. D. (2000). The conceptual complexity of presidents Carter and Clinton: An automated content analysis of temporal stability and source bias. *Political Psychology, 21*, 587-596.

Driver, M. J. (1977). Individual differences as determinants of aggression in the international simulation. In M. G. Hermann (Ed.). *A psychological examination of political leaders* (pp. 337-353). New York: Free Press.

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

Dyson, S. B. (2007). Alliances, domestic politics, and leader psychology: Why did Britain stay out of Vietnam and go into Iraq? *Political Psychology, 28*, 647-666.

Furnham, A. & Schofield, S. (1987). Accepting personality feedback – A review of the Barnum effect. *Current Psychology – Research and Reviews, 6*, 162-178.

George, A. L. (1969). The “operational code”: A neglected approach to the study of political leaders and decision-making. *International Studies Quarterly, 13*, 190-222.

Glass, G. V. & Hopkins, K. D. (1996). *Statistical methods in education and psychology (third edition)*. Boston, London: Allyn and Bacon.

Goldstein, J. S. (1992). A conflict-cooperation scale for WEIS events data. *Journal of Conflict Resolution, 36*, 369–85.

Guttieri, K., Wallace, M. D., & Suedfeld, P. (1995). The integrative complexity of American decision makers in the Cuban missile crisis. *Journal of Conflict Resolution, 39*, 595-621.

Hermann, M. G. (1980). Explaining foreign policy behaviour using the personal characteristics of political leaders. *International Studies Quarterly*, 24, 7–46.

Hermann, M. G. (1987). Assessing the foreign policy role orientations of Sub-Saharan African leaders. In S. G. Walker (Ed.). *Role theory and foreign policy analysis* (pp. 161-198). Durham: Duke University Press.

Hermann, M. G. (2006). Assessing leadership style: Trait analysis. In J. M. Post (Ed). *The psychological assessment of political leaders* (pp. 178-212). Ann Arbor, MI: University of Michigan Press.

Hermann, M. G. and Hagan, J. D. (1998). International decision making: Leadership matters. *Foreign Policy*, 110, 124-137.

Hermann, M. G., Preston, T., Korany, B. and Shaw, T.M. (2001). Who leads matters: The effects of powerful individuals. *International Studies Review*, 3, 83-131.

Holsti, O. (1970). The operational code approach to the study of political leaders: John Foster Dulles' philosophical and instrumental beliefs. *Canadian Journal of Political Sciences*, 3, 123-157.

Keller, J. W. (2005a). Constraint respecters, constraint challengers, and crisis decision making in democracies: A case study analysis of Kennedy versus Reagan. *Political Psychology*, 26, 835-867.

Keller, J. W. (2005b). Leadership style, regime type, and foreign policy crisis behaviour: A contingent monadic peace? *International Studies Quarterly*, 49, 205-231.

Koo, J., Han, J. & Kim, J. (2002). Integrative complexity of South-North Korean correspondences: A time-series analysis, 1984-1994. *Journal of Conflict Resolution*, 46, 286-304.

Leites, N. (1951). *The operational code of the politburo*. New York: McGraw-Hill.

Levi, A. and Tetlock, P.E. (1980). A cognitive analysis of Japan's 1941 decision for war. *Journal of conflict resolution*, 24, 195-211.

Liht, J., Suedfeld, P., & Krawczyk, A. (2005). Integrative complexity in face-to-face negotiations between the Chiapas guerrillas and the Mexican government. *Political Psychology*, 26, 543-552.

Maoz, Z. & Astorino, A. (1992). The cognitive structure of peacemaking: Egypt and Israel, 1970-1978. *Political Psychology*, 13, 647-662.

Maoz, Z. & Shayer, A. (1987). The cognitive structure of peace and war argumentation: Israeli Prime Ministers versus the Knesset. *Political Psychology*, 8, 575-604.

Marfleet, B. G. (2000). The operational code of John F. Kennedy during the Cuban missile crisis: A comparison of public and private rhetoric. *Political Psychology*, 21, 545-558.

McClellan, S. (2008). *What happened: Inside the Bush White House and Washington's culture of deception*. New York: Public Affairs.

McClelland, C. A. (1983). Let the user beware. *International Studies Quarterly*, 27, 169-178.

Mosteller, F. & Tukey, J. W. (1977). *Data analysis and regression*. Reading, MA: Addison-Wesley.

- Moyo, J. (2008, January 12). The 'useful idiot' has gone too far this time. *New Zimbabwe*. Retrieved 3 May 2009, from <http://www.newzimbabwe.com>
- Murray, H. A. (1938). *Explorations in Personality*. New York: Oxford University Press
- Raphael, T. D. (1982). Integrative complexity theory and forecasting international crises. *Journal of Conflict Resolution*, 26, 423-450.
- Satterfield, J. M. (1998). Cognitive-affective states predict military and political aggression and risk taking: A content analysis of Churchill, Hitler, Roosevelt, and Stalin. *The Journal of Conflict Resolution*, 42, 667-690.
- Satterfield, J. M., & Seligman, M. E. P. (1994). Military aggression and risk predicted by explanatory style. *Psychological Science*, 5, 77-82.
- Schafer, M. (2000). Issues in assessing psychological characteristics at a distance: An introduction to the symposium. *Political Psychology*, 21, 511-527.
- Schafer, M. & Walker, S. G. (2006). Democratic leaders and the democratic peace: The operational codes of Tony Blair and Bill Clinton. *International Studies Quarterly*, 50, 561-583.
- Schroder, H.M., Driver, M.J., & Streufert, S. (1967). *Human information processing*. New York: Rinehart & Winston.
- Shannon, V. P. & Keller, J. W. (2007). Leadership style and international norm violation: The case of the Iraq war. *Foreign Policy Analysis*, 3, 79-104.
- Shellman, S. M. (2004). Time series intervals and statistical inference: The effects of temporal aggregation on event data analysis. *Political Analysis*, 12, 97-104.

Smith, A. G. (2008). The implicit motives of terrorist groups: How the needs for affiliation and power translate into death and destruction. *Political Psychology, 29*, 55-75.

Smith, C. P., Atkinson, J. W., McClelland, D. C., & Veroff, J. (Eds.). (1992). *Motivation and personality: Handbook of thematic content analysis* (pp. 393-400). New York: Cambridge University Press.

Suedfeld, P. (1992a). Bilateral relations between countries and the complexity of newspaper editorials. *Political Psychology, 13*, 601-611.

Suedfeld, P., & Bluck, S. (1988). Changes in integrative complexity prior to surprise attacks. *Journal of Conflict Resolution, 32*, 626–635.

Suedfeld, P., Corteen, R. S., & McCormick. (1986). The role of integrative complexity in military leadership: Robert E. Lee and his opponents. *Journal of Applied Social Psychology, 16*, 498:507.

Suedfeld, P., Cross, R.W., & Stewart, M. R. (2008). Individual-level indicators, predictors, and determinants of conflict escalation and de-escalation: A review of the literature. Unpublished manuscript, University of British Columbia.

Suedfeld, P., Guttieri, K., & Tetlock, P. E. (2006). Assessing integrative complexity at a distance: Archival analyses of thinking and decision making. In J. M. Post (Ed). *The psychological assessment of political leaders* (pp. 246-270). Ann Arbor, MI: University of Michigan Press.

Suedfeld, P. & Rank, A. D. (1976). Revolutionary leaders: Long-term success as a function of changes in conceptual complexity. *Journal of Personality and Social Psychology, 34*, 169-178.

Suedfeld, P., & Tetlock, P.E. (1977). Integrative complexity of communications in international crises. *The Journal of Conflict Resolution*, 21, 169-184.

Suedfeld, P., Tetlock, P. E., & Ramirez, C. (1977). War, peace, and integrative complexity: UN speeches on the Middle East problem, 1947-1976. *The Journal of Conflict Resolution*, 21, 427-442.

Suedfeld, P., Tetlock, P.E., & Streufert, S. (1992). Conceptual/integrative complexity. In C.P. Smith, J.W. Atkinson, D.C. McClelland, & J.Veroff (Eds.), *Motivation and personality: Handbook of thematic content analysis* (pp. 393-400). New York: Cambridge University Press.

Tetlock, P. E. (1985). Integrative complexity of American and Soviet foreign policy rhetoric: A time-series analysis. *Journal of Personality and Social Psychology*, 49, 1565-1585.

Thoemmes, F. J. & Conway, L. G. (2007). Integrative complexity of 41 U.S. presidents. *Political Psychology*, 28, 193-226.

Walker, S. G. (2000). Assessing psychological characteristics at a distance: Symposium lessons and future research directions. *Political Psychology*, 21, 597-602.

Walker, S. G., Schafer, M., & Young, M. D. (1998). Systematic procedures for operational code analysis: Measuring and modelling Jimmy Carter's operational code. *International Studies Quarterly*, 42, 175-190.

Walker, S. G. Schafer, M. & Young, M. D. (1999). Presidential operational codes and foreign policy conflicts in the post-cold war world. *Journal of Conflict Resolution*, 43, 610-625.

Walker, S. G., Schafer, M. & Young, M. D. (2006). Profiling the operational codes of political leaders. In J. M. Post (Ed). *The psychological assessment of political leaders* (pp. 215-245). Ann Arbor, MI: University of Michigan Press.

Wallace, M. D., Suedfeld, P. & Thachuk, K. (1993). Political rhetoric of leaders under stress in the gulf crisis. *Journal of Conflict Resolution*, 37, 94-107.

Walker, S. G., & Watson, G. L. (1994). Integrative complexity and British decisions during the Munich and Polish crises. *The Journal of Conflict Resolution*, 38, 3-23.

Winter, D. G. (1973). *The power motive*. New York: Free Press.

Winter, D. G. (1980). An exploratory study of the motives of southern African political leaders measured at a distance. *Political psychology*, 2, 75-85.

Winter, D.G. (1987a). Enhancement of an enemy's power motivation as a dynamic of conflict escalation. *Journal of Personality and Social Psychology*, 52, 41-46.

Winter, D. G. (1987b). Leader appeal, leader performance, and the motive profiles of leaders and followers: a study of American presidents and elections. *Journal of Personality and Social Psychology*, 52, 196-202.

Winter, D. G. (1991c). Measuring personality at a distance: Development of an integrated system for scoring motives in running text. In A. J. Stewart, J. M. Healy, Jr., & D. J. Ozer (Eds.), *Perspectives in personality: Approaches to understanding lives* (pp. 59–89). London: Jessica Kingsley.

Winter, D. G. (1993). Power, affiliation and war: Three tests of a motivational model. *Journal of Personality and Social Psychology*, 65, 532-545.

Winter, D. G. (1996). A motivational analysis of the Clinton first term and the 1996 presidential campaign. *Leadership Quarterly*, 9, 367-376.

Winter, D. G. (2001). Measuring Bush's motives. *ISPP News: International Society of Political Psychology*, 12, 9.

Winter, D.G. (2003a). Asymmetrical Perceptions of Power in Crises: A Comparison of 1914 and the Cuban Missile crisis. *Journal of Peace Research*, 40, 251-270.

Winter, D.G. (2003b). Personality and political behaviour. In D. O. Sears, L. Huddy & R. Jervis (Eds.). *Oxford handbook of political psychology*. New York: Oxford University Press.

Winter, D. G. (2005). Things I've learned about personality from studying political leaders at a distance. *Journal of Personality*, 73, 557-584.

Winter, D. G. (2006). Measuring the motives of political actors at a distance. In J. M. Post (Ed). *The psychological assessment of political leaders (pp. 153-177)*. Ann Arbor, MI: University of Michigan Press.

Winter, D.G. (2007). The role of motivation, responsibility, and integrative complexity in crisis escalation: Comparative studies of war and peace crises. *Journal of Personality and Social Psychology*, 92, 920-937.

Winter, D. G., Hermann, M. G., Weintraub, W., & Walker, S. G. (1991a). The personalities of Bush and Gorbachev at a distance: Procedures, portraits and policy. *Political Psychology*, 12, 215-245.

Winter, D. G., Hermann, M. G., Weintraub, W., & Walker, S. G. (1991b). The personalities of Bush and Gorbachev at a distance: Follow-up on predictions. *Political Psychology*, 12, 457-464.

Winter, D. G. & Stewart, A. J. (1977). Content analysis as a method of studying political leaders. In M. G. Hermann (Ed.), *A psychological examination of political leaders (pp. 27-61)*. New York: Free Press.

Young, M. D. (2000). Automating assessment at a distance. *Political Psychologist*, 5, 17–23.