

# European Union Politics

<http://eup.sagepub.com/>

---

## **Decision-Making Models, Rigor and New Puzzles**

Bruce Bueno de Mesquita

*European Union Politics* 2004 5: 125

DOI: 10.1177/1465116504040448

The online version of this article can be found at:

<http://eup.sagepub.com/content/5/1/125>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *European Union Politics* can be found at:**

**Email Alerts:** <http://eup.sagepub.com/cgi/alerts>

**Subscriptions:** <http://eup.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://eup.sagepub.com/content/5/1/125.refs.html>

>> [Version of Record](#) - Mar 1, 2004

[What is This?](#)



**European Union Politics**

DOI: 10.1177/1465116504040448

Volume 5 (1): 125–138

Copyright© 2004

SAGE Publications

London, Thousand Oaks CA,

New Delhi

## Forum Section I

# Decision-Making Models, Rigor and New Puzzles

◆ **Bruce Bueno de Mesquita**

*New York University and Stanford University, USA*

### ABSTRACT

Comparative tests of predictive accuracy across models are exceptionally rare in political science. The collection of articles in this issue provides a rigorous, systematic evaluation of alternative models for explaining and predicting decision-making within the European Union. I examine how alternative models were evaluated and raise questions about the differences in the extent to which the operational definitions of variables match contending theoretical approaches. I also raise questions regarding the difference between models with regard to whether the issues are part of a repeated game or a single-shot game. Finally, I propose future tests to hone in on some of the puzzles raised by the findings reported here.

### KEY WORDS

- models
- measurement
- prediction
- post-diction
- reliability

Our understanding of political decision-making is most likely to progress by combining rigorous, explicit theory with equally rigorous empirical tests. In this way we can gradually sort out the circumstances in which one type of model is most effective and other circumstances for which some other model proves better suited. With hard work and a bit of luck, we may even gradually develop covering laws and associated theory that encompasses seemingly discrete approaches into a unifying account. The studies in this issue and the accompanying edited volume (Thomson et al., forthcoming) are directed toward advancing these goals. Indeed, this is a remarkable collection of essays precisely because the essays are unified in purpose and unified in the data used to evaluate European Union decisions across 162 controversial issues. What is more, these studies represent the most rigorous assessment undertaken to date of alternative explanations of EU decision-making. The model testing performed here greatly advances our understanding while, at the same time, compelling us to confront difficult issues regarding testing and model assessment. In this article, I address the model assessment issues that I consider most important. Naturally this will lead me to be critical of some of the procedures used here and of the inferences associated with those procedures. But I should not be misunderstood. Although my task is to bring a critical eye to bear, the studies in this issue are, in my view, the finest work done thus far in applying rigorous standards to the empirical evaluation of competing explanations of decision-making, and they convincingly demonstrate fundamental insights into how the European Union works.

The models and algorithms – for not all procedures for predicting decisions involve deductive models – presented here represent a wide range of approaches. Some, such as the exchange model and the expected utility model – referred to as the challenge model to avoid a confusing proliferation of meanings for ‘EU’ – focus on bargaining. Others investigate procedures or the impact of formal rules on outcomes. Many of the models are presented within a cooperative game theory framework, although recognizing that these are essentially a reduced form of some (perhaps unspecified) non-cooperative game. Others are more explicitly embedded in a non-cooperative game setting. The studies further divide along another important dimension. Some provide a predictive account of the process that leads to outcomes, whereas others eschew any effort to account for the process, providing only a prediction about the final policy choice on each issue. In other words, the studies here represent a diverse array of approaches explaining or at least predicting EU decisions.

Several essays are specifically concerned with comparing results from various models with one another. Achen (forthcoming b), for instance, concludes that the baseline predictions (i.e. the unweighted median and

mean) perform best, followed by the compromise, exchange and coalition models. These are followed in accuracy by the weighted median and the challenge model. According to Achen, all the other models fall into a group that does significantly worse. This assessment is itself a bit confusing. Arregui et al. (in this issue) report no statistical difference between the mean absolute error for the exchange model and the challenge model. This point, however, is less important than the process by which testing has been done here. That process fosters new insights and introduces puzzles for future research. It is to these matters that I now turn.

### **A risk with post-diction**

The contributors to Thomson et al. (forthcoming) have made an admirable attempt to ensure comparability in tests. This has been pursued by making sure that everyone worked from the same database so that results could be readily compared. However, questions arise in treating the test results as completely comparable across models and issues or as easily generalized to situations of true prediction; that is, the application of contending models and algorithms to issues with as-yet-unknown outcomes. I say this for several reasons.

How does post-diction differ from prediction? Of course, the most obvious difference is testing against known outcomes versus testing against unknown outcomes. Ordinarily we are tempted to think that it is easier to account for a known outcome than to predict one that is unknown. With known outcomes, data are expected, perhaps misleadingly, to be of higher quality, any exogenous shocks are known about at the time of testing and so corrections can be made for them, and so forth. In that sense, we generally expect a decline in 'goodness of fit' as we move from in-sample analyses to out-of-sample or genuinely predictive assessments. Yet there also are some questions about drawing predictive inferences about model reliability from ex post analyses.

One way to evaluate the difference between post-diction and prediction is to take advantage of information about performance in ex ante and ex post settings. Several of the models in Thomson et al. (forthcoming) have previously been subjected to large numbers of tests 'in real time'. That is, they are regularly applied to problems for which the outcome was unknown at the time of the investigation. This is true, for instance, of the exchange model, the compromise model and the challenge model. These are, incidentally, the models in Thomson et al. (forthcoming) that are concerned with bargaining and with the process (as distinct from the procedures) by which policy

positions evolve, leading eventually to a final choice. These were, as well, among the better-performing models in this investigation, but their performance here appears significantly worse than is true of their externally audited performance in genuinely predictive applications. Stanley Feder (1995), for instance, reports that the challenge model – he refers to it as the expected utility model – proved accurate about 90% of the time in a truly predictive study conducted by the US government. He also reports that, when the model's predictions disagreed with the predictions of the experts whose data were used to generate the model's results, the expected utility model almost always proved right and the expert predictions proved incorrect. We can see similar patterns in the ex post evaluation of ex ante predictions published in the academic literature that are based on this model (a sample includes Bueno de Mesquita, 1984, 2002; Ray and Russett, 1996; James and Lusztig, 1997; Kugler and Feng, 1997). Real-time applications of the exchange and compromise models by Stokman and others can likewise boast a record that appears even stronger than the one achieved in the tests presented in the volume edited by Thomson and his colleagues.<sup>1</sup> This leaves us with a puzzle. Why would ex ante predictions appear to have a better track record than ex post predictions?

In trying to answer this question, I consider the general statistical findings across the models tested here. Two of the best predictors proved to be the baseline mean and the unweighted median voter positions. This is somewhat surprising. These two predictions are the product of simple algorithms. Little, if any, theory guides the expectation that the mean or unweighted median voter position is favored. What is more, these two baseline predictors involve no process of bargaining or procedures that induce equilibrium. These predictions simply take stakeholder positions on each issue as given and then pick a winning position. In these approaches, stakeholders have no opportunity to switch positions, forge compromises, reshape coalitions, trade with one another, etc. Yet the data are believed to reflect the *initial* position of each stakeholder at the start of the decision-making process and not the position at the time a decision was taken. There is no danger that this belief is mistaken for the approximately one half of issues analyzed for which final positions and initial positions remained the same. But for the half that seem to have involved a fair amount of horse-trading or compromise, we are left with a puzzle.

Perhaps the common belief that compromises are routinely forged through bargaining, leading to shifting positions, is mistaken. After all, half of the cases here show no positional shifts. But then half do. Perhaps the experts mistook initial policy positions for ideal points or for final positions in those cases where there is no recorded shift in position from the beginning

to the resolution of an issue, a matter to which I return in a later section. It is noteworthy that Stokman and Thomson (in this issue) indicate that the experts could not distinguish empirically between actors' preferences (i.e. ideal points) and their initial positions. If the input data on positions reflect the experts' views of initial positions tainted by ex post knowledge of where each stakeholder ended up, then models that correctly predict a dynamic process will prove less accurate in post-dictive studies than in predictive investigations, because the input data will misrepresent the stage at which the model enters the analytic process. This is thus far an unexamined potential empirical disadvantage arising from post-diction as compared with prediction.

About half of the issues examined here are based on data taken after the outcome was known, and the other half were collected during the decision-making process. It would be a useful addendum in future research to know whether there are systematic differences in how well contending models performed across these two distinct sets of issues. There is tantalizing evidence, albeit based on a small sample, from issues data collected by Maurice Rojer (1999) that there was not a significant difference in model performance as a function of whether the data were collected ex post or during the process. If this finding holds up when applied to the full data set and to all of the models, it will provide an additional and highly important source of further confidence in the general reliability of the findings reported here for future studies of EU decisions. It will also facilitate the choice of models with which to predict decisions based on repeated interaction in the EU, a topic returned to later in this article.

In addition to a statistical assessment of any differences in performance of the contending models based on whether data were ex ante or ex post, future studies might also be specifically designed to draw out the importance of real-time prediction. Perhaps this or some other team of scholars will follow up with a study of a large number of issues that are chosen prior to their resolution, whether in the European Union (with its repeated play) or in some other political settings, including those without repeated interactions or with stochastic but not repeated play. Then there will be no possibility that the data are tainted by reconstructions of history by experts who know the outcome.

Bueno de Mesquita et al. (2001) undertook such a study of the Good Friday Agreement in Northern Ireland. They investigated 11 specific issues regarding implementation of the terms of the agreement and they fixed a time frame – through the year 2000 – during which their predictions were to apply. This was the time frame specified within the Good Friday Agreement. They then compared the challenge model with prospect theory predictions – submitting to the journal that ultimately published their analysis the

predicted results more than a year before the period for which results were predicted – and used detailed news accounts of what transpired to assess what the actual outcomes on implementation were as compared with those predicted by the contending models. Such a truly predictive exercise on a larger scale would be a useful addendum to the studies discussed here and would provide a direct means to assess the extent to which an expert's ex post knowledge influences his or her presumptively ex ante data inputs.

### **Bridging theory and data**

Thomson and Hosli (forthcoming) make the point that the models examined in Thomson et al. rely on 'similar types of data'. Little more is said on this subject. However, there are questions regarding the comparability of the linkage between the models, data and tests. I touch on these only briefly because, although the issues raised are real and important, they are unlikely to have had a significant impact on results.

### **Policy position data**

Consider the data on policy positions. Many of the models assume that the data on policy positions reflect stakeholders' ideal points. Stokman and Thomson (this issue) make clear that the positional data were not designed to measure ideal points. They indicate that the experts were asked to 'indicate the policy alternative *initially* favoured by each stakeholder after the introduction of the proposal before the Council formulated its common position'. This definition is well suited to the exchange, challenge and compromise models; it is very close to the positional information those models logically require. But the other models in Thomson et al. (forthcoming) are logically grounded in knowing the ideal point or bliss point of each stakeholder. This is problematic empirically for at least two reasons.

My view regarding ideal points differs significantly from that expressed by Achen (forthcoming a). I believe that knowing ideal points is nearly impossible, certainly ex ante and quite possibly ex post. Stated, observable bargaining positions are presumably strategically chosen. Stakeholders, especially in a repeated game, may have few reasons to reveal their ideal points. There may be strategic advantages for some stakeholders in signaling that they are relatively harder-line or softer-line than they truly are in order to extract better concessions from others. Uncertainty about one's ideal point can certainly prove advantageous in bargaining, as demonstrated in much of the literature on incomplete information games. The challenge model

explicitly rules out inputs based on policy ideal points. The stated (initial) bargaining position for each stakeholder is used to impute the shape of an indifference curve in which stakeholders trade off getting the policy outcome they desire against getting political credit for their role in helping (or hindering) the construction of an agreement. A decision-maker's ideal point for an issue is assumed to fall at the extreme end of an indifference curve. Consequently, the marginal rate of substitution (MRS) between policy goals and credit-seeking is radically different from the MRS just about anywhere else on the indifference curve. This means that ideal policy point data would provide a distorted view of the bargaining process as compared with how the MRS looks at the policy position the stakeholder has chosen to reveal publicly. Since the data used here are not positional ideal points, the challenge model is relatively privileged compared with models that assume bliss point information.

Models that assume that policy ideal points are critical in shaping choices (as opposed, for instance, to models concerned with a policy-credit contour) are inherently disadvantaged in empirical applications because of the difficulty in learning true ideal points. Perhaps it would be valuable for those who use the coalition model or other procedural models that rely on ideal points to work through the strategic elements that may allow the translation of observed positions into estimates of the feasible range of ideal points for individual decision-makers. This is a difficult task, but without it these models – however informative they are on a theoretical level – are unlikely to perform well in empirical settings. Indeed, they generally do not fare well in Thomson et al. (forthcoming). I return to this point when I discuss repeated play and then suggest one basis on which we may infer conditions under which ideal point data are less useful than observed policy position.

### **Salience data**

Salience means at least two, and possibly three, different things in the models in Thomson et al. (forthcoming). For some models, its relevance is tied to the Nash bargaining solution. In that sense, salience dictates the curvature of a stakeholder's indifference curve between outcomes on alternative issues (which should not be confused with policy and credit outcomes on a single issue). It plays just such a role in Achen's (forthcoming a) formulation and, in part, approximates that role in the exchange model. Salience, however, has a different meaning in the compromise, challenge and exchange models. In those models, salience measures the priorities of each stakeholder across issues. The ideas of salience as curvature and of salience as priority are not reflected equally in the data collection. That is a natural consequence of a



pioneering study faced with a daunting and unprecedented data collection task. In all likelihood the differences in definition are small enough that they do not explain more than a small part of the difference in how well each model has performed. The difference may be especially consequential for the procedural models.

### Repeated games and single-shot games

The European Union offers an empirically rich environment to test models of decision-making. Inherent characteristics of that environment also may favor the selection of cooperative equilibria over non-cooperative equilibria. Credible commitments are more easily made between players engaged in repeated interactions than is true for players engaged in a single episode of interaction. Cooperative games make modeling side payments easy by assuming away the possibility that a player will renege. In essence, many of the issues surrounding time inconsistency are obviated by a cooperative game structure. In a non-cooperative game setting (recalling that cooperative games are reduced forms of non-cooperative games, but not all non-cooperative games are sensibly reduced to a cooperative game), repeated play provides opportunities for punishment strategies that can produce cooperation and compromise where that would be difficult if the players did not repeatedly interact. One interesting observation in this regard is that cooperative models fare better in Thomson et al. (forthcoming) than do non-cooperative models. But then European Union members can anticipate repeated interactions.

The expectation of indefinite repetition facilitates logrolls within and across issue areas. If a pair of states agrees to a positional trade and one reneges, there are ample opportunities on future issues for the aggrieved party to punish the party who reneged. Given the credibility of such a threat of future punishment, rational expectations lead to few instances of cheating on agreements and, therefore, to improved predictive success with reduced-form cooperative games. Interestingly, the challenge model's track record in *ex ante* predictions is based overwhelmingly on studies of non-recurrent interactions (Feder, 1995, 2002; McGurn, 1996; Ray and Russett, 1996). The procedural models rely on repetition within known, existing decisional structures to induce equilibria while avoiding the McKelvey/Schofield chaos results (McKelvey, 1976, 1979; Schofield, 1978). The logrolling models rely implicitly on repeated play to ensure that commitments are credible. Otherwise, these models run into the problem that every trade has a counter-trade, provided that no procedural rules preclude such opportunities.

Repeated play establishes an important rational foundation for the limited performance of the procedural models represented in Thomson et al.

(forthcoming). Procedural models are concerned with how the formal rules of decision-making contribute to shaping outcomes. They ignore informal norms of behavior that emerge as properties of repeated-play equilibria, except to the extent that the formal rules are specifically designed to anticipate and create those norms. Cooperative models are best suited to capturing the dynamics of strategic interactions in the EU's repeated-play environment. In this regard, I repeat an argument I made in Bueno de Mesquita and Stokman (1994: 73–4):

The member states of the European Community presumably value the EC as an institution. Otherwise it would cease to exist. If they fail to reach compromise settlements on issues requiring unanimity, then the failure may threaten the very survival of the Community.

When failure to agree endangers the integrity of the EU, decision-makers face an elevated and cumulative cost if they fail to compromise, making cooperation more likely. Indeed, this may be the major factor accounting for the superior performance of the exchange and compromise models as compared with the challenge or procedural models when analyzing positional shifts between the initiation and termination of issues, as demonstrated by Arregui et al. in their article in this issue on shifts in policy positions. There is considerable evidence that the European stakeholders and institutions try to achieve unanimity *even when not required by the procedures*. Thus, procedures are not permitted to dictate the action when strict adherence to them endangers the institution. Norms of trust and cooperation naturally emerge because EU participants know that, in addition to the legalistic procedural constraints, they must interact with one another in the future and therefore stand to gain by learning to cooperate and compromise.

## Defining and measuring error

Knowing when a model's predictions are right or wrong, or how right or wrong they are, is a difficult matter. Thomson et al. (forthcoming) is an exemplar of excellent efforts to cope with the difficulties inherent in measuring error. Although this is unlikely to alter any of the conclusions, there are nevertheless two issues that I believe should be examined so that future undertakings may, at least, consider the rationale for alternative tests.

### Mean absolute error

One of the prominent tests used throughout Thomson et al. (forthcoming) to assess accuracy is the mean absolute error (MAE). This test is simple and

seems easily interpreted. With the issue continua normalized from 0 to 100, this test also appears to facilitate comparisons across issues and across models. It simply takes the absolute value of the difference between a predicted outcome and the observed outcome on each issue for a given model and divides that value by 100, the range of the issue continua and, therefore, the presumptive maximum possible predictive error. The average across issues is then calculated for each model, and that is the reported MAE. This is a sensible test but it can be misleading in ways that are easily addressed by an additional, alternative specification.

Consider the following. Suppose the predicted outcome on an issue is 50 and the observed outcome is 0. The worst possible predicted outcome would have been 100, and the MAE reports that the error in this case is 50%. But what if the predicted outcome is 0 and the observed outcome is 50? Again the MAE would report a 50% error, but are the errors truly the same in these two cases? If the observed outcome is 50, then the worst that one could do is to predict either 0 or 100. In either case, the difference between the observed and predicted outcome is 50, as was true in the previous example. However, in this example it is not possible to make a mistake of 100 points in the prediction. Since the largest possible error is 50 points, the prediction of 0 (or 100) yields an error of 100% of what was feasible. Whenever the observed outcome is in the interior of the issue continuum, the MAE *understates* the predictive error for a model, at least to the extent that the maximum feasible error is a relevant consideration.<sup>2</sup>

### Dichotomous choices

Many of the issues in the data set used for Thomson et al. (forthcoming) involved dichotomous choices. Indeed, many decisions in politics ultimately involve saying yes or no, voting for or against a particular bill, etc. When a model is designed to use positional data that can be located anywhere on a continuum but the ultimate decision is a binary choice such as yes or no (or a ranked choice, with infeasible alternatives between rankings), then we must think about what a prediction on the interior of the continuum (or between ranked positions) means. The approach taken in this issue is to treat the interior prediction as suffering from an error that can be calculated according to the MAE or related indicators. Yet we know that, for many such issues, the interior values are not in the feasible set and so we are testing against irrelevant alternatives. It seems to me that predictions in these cases have a perfectly straightforward point interpretation, given that the models tested here assume that preferences are single peaked. A stakeholder faces a binary choice between voting 0 or 100, but has an intensity of preference for the

outcome that can vary, in its normalized form, between 0 and 100. So, a score of 50 seems to indicate indifference between an outcome of 0 and an outcome of 100. A score of 30 seems to indicate that such a stakeholder will vote 0, albeit with less enthusiasm than someone whose position is, say, 10 on the scale. Likewise, someone with a score of 80 can be expected to vote for the outcome at 100 because it is the feasible action closest to the stakeholder's preference.

These inferences follow directly from the assumption that stakeholder preferences are single peaked. Otherwise, none of the models studied here could be used to draw inferences about how stakeholders will respond to alternatives located away from their own position. If the preferences are single peaked and the choices are binary (or ranked), then it seems unnatural to interpret a predicted position of 30 and an outcome of 0 as involving a 30% error. When the true choice is binary, the outcome is 0 and the predicted position is less than 50, then the predicted action is the action at 0. Likewise, if the outcome is 100, the choice is binary and a stakeholder's position is above 50, the predicted action is 100 and there is no error in the prediction. If the position were below 50, then the predicted action would be 0 and, if the outcome were 100, the error would be 100%. Comparable inferences can be drawn regarding predictions when data are ranked according to discrete values but positions are plotted continuously along a line segment.

If the procedure to measure predictive error on binary choices takes the assumption of single-peaked preferences and the feasible set seriously, then the MAE procedure can seriously mislead, especially if some models favor interior positions on the continuum whereas others favor boundary predictions. Consider the set of 'predicted outcomes' from three hypothetical models applied to binary choices as displayed in Table 1. The MAE for the first model is 45%, for the second model it is 40%, and for the third model it is 30%. One might well infer, then, that model 1 performs worst, that model 2 is second worst, and that model 3 performs best. However, knowing that the choice is binary, so that only two actions are permitted, and that preferences are single peaked, and assuming reliable data, we predict 100% of the outcome actions correctly with the first model, 80% correctly with the second model, and 70% correctly with the third model. Given this risk of reversal in the judgment about how well models perform, we should want to consider more closely the extent to which specific tests of goodness of fit comport with the theoretical assumptions underlying the models and the data. It may make no difference to the results for the set of models tested by Thomson et al. (I do not have the data to ascertain whether it does or does not), but it is a reminder that seemingly straightforward tests can have significant problems embedded in them if we do not match the test structure to the assumptions behind the data.

**Table 1** Measuring error for binary choices, assuming single peaked preferences

<i>Outcome</i>	<i>Predicted outcomes</i>		
	<i>Model 1</i>	<i>Model 2</i>	<i>Model 3</i>
0	45	60	0
0	45	20	0
0	45	40	0
0	45	40	100
0	45	40	100
100	55	40	100
100	55	80	100
100	55	60	100
100	55	60	100
100	55	60	0

## Conclusion

The volume edited by Thomson et al. (forthcoming) is an outstanding example of rigorous theorizing and testing of models concerned with decision-making. Future research might further improve our knowledge of such models and of the decision-making process by amending the testing process in a few ways. First, I recommend selecting a large number of as yet unresolved issues so that there is no risk that expert data are tainted by ex post knowledge of how things turned out. Second, while applying data on the same set of issues to each model, I suggest collecting data in such a way that semantic similarities in variable labels do not cloud differences in the precise meaning each model attaches to the variable. For instance, positional data should be collected three ways – probably from distinct groups of experts – to reflect ideal point information; initial position information; and current stated position at the time the experts are interviewed. Testing for accuracy should be linked as closely as possible to the underlying assumptions in the data. For instance, when position data are continuous but the outcome is a binary or ranked choice, the assumption of single-peakedness and restrictions on the feasible set should dictate that predictions be translated into the feasible action they indicate rather than being treated as if single-peakedness and restrictions on the feasible set were not relevant. Finally, tests in repeated-play settings such as the EU and in non-repeated-play settings can help further to sort out which models work best in which settings. The research in Thomson et al. (forthcoming) focuses our attention on the importance that

repeated play, stochastic games and single-shot games may have for model selection and for accurate explanations of the decision process inside and outside the EU.

It is the nature of normal science that progress – and Thomson et al. most assuredly demonstrate significant progress in our understanding of decision-making inside and outside the EU – inevitably identifies new puzzles to be solved. That is true here. We can look forward to future efforts by this or other teams of researchers to advance our understanding further and to extend the frontiers of knowledge and of new puzzles to be investigated.

## Notes

---

- 1 This is documented at <http://www.decide.nl/index2E.html>.
- 2 In private correspondence Frans Stokman informed me that he and his colleagues conducted the test suggested here. As anticipated, it did not alter any conclusions and so, to conserve space, they did not report the results in the text.

## References

---

- Achen, Christopher (forthcoming a) 'Institutional Realism and Bargaining Models', in Robert Thomson, Frans Stokman, Christopher Achen and Thomas König (eds) *The European Union Decides*.
- Achen, Christopher (forthcoming b) 'Evaluating Political Decision-Making Models', in Robert Thomson, Frans Stokman, Christopher Achen and Thomas König (eds) *The European Union Decides*.
- Arregui, Javier, Frans N. Stokman and Robert Thomson (forthcoming) 'Compromise, Exchange and Challenge in the European Union', in Robert Thomson, Frans Stokman, Christopher Achen and Thomas König (eds) *The European Union Decides*.
- Bueno de Mesquita, Bruce (1984) 'Forecasting Policy Decisions: An Expected Utility Approach to Post-Khomeini Iran', *PS* 18(2): 226–36.
- Bueno de Mesquita, Bruce (2002) *Predicting Politics*. Columbus: Ohio State University Press.
- Bueno de Mesquita, Bruce and Frans Stokman (eds) (1994) *European Community Decision Making: Models, Applications, and Comparisons*. New Haven, CT: Yale University Press.
- Bueno de Mesquita, Bruce, Rose McDermott and Emily Cope (2001) 'The Expected Prospects for Peace in Northern Ireland', *International Interactions* 27(2): 129–67.
- Feder, Stanley A. (1995) 'Factions and Policon: New Ways to Analyze Politics', in H. Bradford Westerfield (ed.) *Inside CIA's Private World: Declassified Articles from the Agency's Internal Journal, 1955–1992*. New Haven, CT: Yale University Press.
- Feder, Stanley A. (2002) 'Forecasting for Policy Making in the Post-Cold War Period', *Annual Review of Political Science* 5: 111–25.

- James, Patrick and Michael Lusztig (1997) 'Assessing the Reliability of Predictions on the Future of Quebec', *Quebec Studies* 24: 197–209.
- Kugler, Jacek and Yi Feng (eds) (1997) 'The Expected Utility Approach to Policy Decision Making', *International Interactions* 23(3–4).
- McGurn, William (1996) 'We Warned You: Hong Kong Is Losing Its Edge', *Far Eastern Economic Review*, 13 June.
- McKelvey, Richard (1976) 'Intransitivities in Multidimensional Voting Models and Some Implications for Agenda Control', *Journal of Economic Theory* 12: 472–82.
- McKelvey, Richard (1979) 'General Conditions for Global Intransitivities in Formal Voting Models', *Econometrica* 47: 1085–112.
- Ray, James Lee and Bruce Russett (1996) 'The Future as Arbiter of Theoretical Controversies: Predictions, Explanations and the End of the Cold War', *British Journal of Political Science* 26: 441–70.
- Rojer, Maurice (1999) 'Collective Decision-Making Models Applied to Labour Negotiations in the Netherlands: A Comparison between an Exchange Model and a Conflict Model', *Rationality and Society* 11: 207–35.
- Schofield, Norman (1978) 'Instability of Simple Dynamic Games', *Review of Economic Studies* 45: 575–94.
- Thomson, Robert and Madeleine O. Hosli (forthcoming) 'Explaining Legislative Decision Making in the European Union', in Robert Thomson, Frans Stokman, Christopher Achen and Thomas König (eds) *The European Union Decides*.
- Thomson, Robert, Frans N. Stokman, Christopher H. Achen and Thomas König (eds) (forthcoming) *The European Union Decides*.

---

## About the author

**Bruce Bueno de Mesquita** is Silver Professor and Chair, Department of Politics, New York University, 726 Broadway, 7th Floor, New York, NY 10003, USA; he is also a Senior Fellow at the Hoover Institution, Stanford University, Stanford, CA 94305–6010, USA.

Fax: +1 650 723 1687

E-mail: [bdm@hoover.stanford.edu](mailto:bdm@hoover.stanford.edu)

---