

Jacques Hymans
(hymans1@fas.harvard.edu)
Presentation prepared for Gov 3009: research workshop in applied statistics
DRAFT: DO NOT CITE
4/22/01

Measuring the unmeasurable: national identity as a variable

I. Introduction

In my dissertation, I attempt to explain why some states acquire nuclear weapons, while others abstain from doing so. The argument focuses on the preferences of leaders at the utter pinnacle of state power. I argue that different "national geopolitical identities" (NGIs) produce different cognitive and emotional predispositions on the bomb and other related nuclear policy matters. In particular, leaders whose NGIs can be characterized as "oppositional nationalist"—national self-definition as opposition to a perceived menacing enemy combined with confidence in the nation's ability to face the enemy down—are likely both to want and to decide to acquire nuclear weapons. By contrast, leaders whose NGIs are of other types are unlikely either to want or to decide to "go nuclear."¹

This brief statement of the theoretical argument indicates the difficulty of theory-testing in this case. The problems are daunting even *before* we consider testing the posited relationship between the variables. These more "typical" issues of social science research methodology can only be tackled after establishing the validity and reliability of measurement of *each variable on its own*.² On the independent variable side, national identity has often been considered so diffuse a variable as to be essentially unmeasurable, that is if it exists at all.³ On the dependent variable side, preferences are notoriously difficult to ascertain, and this difficulty is increased markedly in an issue-area where debates and decisions happen in the deepest secrecy.⁴

In this paper, I aim to establish the validity and reliability of the measurement strategies I have adopted for my independent variable, "national geopolitical identity." In order to do so, I first outline a set of methodological principles and their generic practical consequences that I have culled from the literature on validity, and I argue for a set of specific steps I have taken in order to arrive at a valid measure of the identity variable. I then consider the topic of reliability,

¹ I argue that institutional, political, and psychological factors particular to the choice for nuclear weapons make it one of those "loneliest decisions a person can ever make," as the cliché goes. In other words, these factors generally allow the leaders' preferences to be translated into policy directly, without significant political debate within or outside the state. By contrast, policy decisions on other related issues, such as whether or not to acquire nuclear technology or to sign the non-proliferation treaty, are more likely to depend on the interaction of the preferences of various actors, as well as on the political weight that their institutional position gives them.

² Adcock and Collier argue that research in political science should spend much more time on this prior step than it typically does. Robert Adcock and David Collier, "Connecting Ideas with Facts: The Validity of Measurement, with Examples from Comparative Research on Democracy," presented at the APSA Annual Meeting, Washington DC, August 31, 2000.

³ Consciousness is one example of a concept that we accept exists (at least in humans) even though it is essentially impossible to measure. Indeed, the difficulty of pinning down national identity can be understood by the fact that one often used synonym for it is "national consciousness."

⁴ One might add that given the moral weightiness of the issue, it might be the case that individuals would not speak freely even in those secret meetings. Eli Levite has mentioned to me that Moshe Dayan once said that some ideas are so terrible that he dare not admit them even to himself.

including presenting the results of a test of inter-coder reliability of the measures I have developed.

II. Three Types of Validity

In a general sense, validity can be defined as "measuring what we think we are measuring."⁵ This actually involves two steps: first, defining clearly the unobservable variables on the conceptual plane, and second, developing clear arguments for why our "instruments" are appropriate for measuring those variables empirically. Recognizing this difference between the "conceptual" and the "operational" levels of validity is particularly important when the underlying variable we are attempting to measure is not directly observable—as for instance when we are making claims about attitudes, beliefs, emotions, and other internal states including identities.⁶ In such cases, it is even more important than usual to have an "auxiliary measurement theory," or an explicit set of assumptions about the linkages between measured and unmeasured variables.⁷

Each step in the validation process—the conceptual and the operational—requires separate explorations of different aspects of validity. Indeed, there is a welter of meanings associated with the term "validity," and I have condensed the literature on this subject to those points that seem most relevant on a practical level.⁸ Putting first things first, I am leaving out any discussion of the validity of claimed correlations between variables, in favor of concentrating solely on the validity of the identity variable on its own.⁹ Three aspects of validity appear central on both the conceptual and operational levels: "correspondence," "differentiation," and "robustness."¹⁰

⁵ Gary King, Robert O. Keohane, and Sidney Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton, NJ: Princeton University Press, 1994), p. 25.

⁶ Stuart A. Rice, "Objective indicators of subjective variables," in Paul F. Lazarsfeld and Morris Rosenberg, eds., *The Language of Social Research* (New York: The Free Press, 1955), pp. 35-37.

⁷ H. M. Blalock, Jr., "Auxiliary measurement theories revisited," in J. J. Hox and J. de Jong-Gierveld, eds., *Operationalization and Research Strategy* (Amsterdam: Swets and Zeitlinger, 1990), pp. 33-48. Auxiliary measurement theories are unusual kinds of "theories" in that they are not really testable; in order to test them one would need a way directly to measure the underlying variable, yet the lack of such a direct measure is what led to the need for the auxiliary measurement theory in the first place. But I will argue that there are a number of ways to maximize their plausibility.

⁸ For an extensive list, including a call for still more types of validity, see David Brinberg and Joseph E. McGrath, "A network of validity concepts within the research process," in David Brinberg and Louise H. Kidder, eds., *Forms of Validity in Research* (San Francisco: Jossey-Bass, 1982), pp. 5-22.

⁹ In particular, I refrain from discussing two of the most frequently mentioned sub-types of validity, "internal" and "external" validity, which are primarily geared toward examining conclusions about the relationships between variables. "Internal" and "external" validity are described in Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1963).

¹⁰ The terms "correspondence" and "robustness" are borrowed from *Ibid.*, p. 12, but the meanings have been somewhat altered in light of other writings on validity that I will mention. The term "differentiation" is my own, but it is parallel to what is called "discriminant validation." Donald T. Campbell and Donald W. Fiske, "Convergent and discriminant validation by the multitrait-multimethod matrix," *Psychological Bulletin*, March 1959, pp. 81-104, cited in Herbert B. Asher, "The Research Process," in Herbert B. Asher et al., *Theory-Building and Data Analysis in the Social Sciences* (Knoxville: University of Tennessee Press, 1984), p. 14.

III. Correspondence

Correspondence on the conceptual level

On the conceptual level, correspondence has to do with the "fit" between the proposed scientific definition of the term and how the term is understood in ordinary language. "Power," "beliefs," and "democracy" come with meanings attached to them and we risk confusion if we define them without any regard to those meanings. Scientific terms in the social sciences are almost always attempts to define more precisely the various dimensions of concepts and "variables" that we often employ in everyday speech about our individual and collective lives. Establishing a tight fit between ordinary and scientific language is not only about preventing confusion, however. It is also a first, albeit questionable means of ensuring that the concept is pointing to some actual social fact.¹¹ This is important, for it is fruitless to try to measure something that does not exist.

James Fearon has made the most serious effort to date to develop a definition of identity that is both faithful to ordinary usage and susceptible to social-scientific measurement.¹² Fearon distinguishes between two types of identity: "personal" identity held by individuals, and "social" identity held by collectivities. This terminology is not terribly felicitous, because it implies that national identities must be identities held by nations as a whole. In fact, a careful reading of Fearon indicates that national and other group identities can be conceived of either as "social," discursive structures or as "personal," cognitive-affective complexes that are experienced as intensely and defined as individualistically as a person's "individual" identity.¹³ It is in this latter sense that I employ the term "national identity." In defining it, I therefore start from Fearon's definition of "personal" identity: "a set of attributes, beliefs, desires, or principles of action that a person thinks distinguish her in socially relevant ways and that a) the person takes a special pride in; b) the person takes no special pride in, but which so orient her behavior that she would be at a loss about how to act and what to do without them; or c) the person feels she could not change even if she wanted to."¹⁴ My somewhat modified and condensed version of this definition, adapted to the concept of "national identity," goes as follows: "*National identity is the set of people, values, interests, and capabilities by which a person consciously or unconsciously defines her national group and about which she feels some level of pride or shame.*" There are four key changes that my definition makes to Fearon's base definition. First is the inclusion of a reference to (human) boundaries—not necessary in defining "individual" identity but necessary in defining "group" identity.¹⁵ Second is the set "values, interests, and capabilities" instead of "attributes, beliefs, desires, and principles of action." I would argue that this is largely a matter of selecting terminology that is more commonly applied and more precisely suited to the national identity case; one can equate values with beliefs, interests with desires, and capabilities with

¹¹ I am not arguing that everything we refer to in ordinary language in fact exists, but I am arguing that in the absence of any direct observability in a scientific sense, this "click of recognition" counts for something. This "click of recognition" is sometimes called "face validity," though that term is also used in other ways. See Louise H. Kidder, "Face validity from multiple perspectives," in Brinberg and Kidder, eds., *New Directions for Methodology*, op. cit., pp. 41-58.

¹² James Fearon, "What is Identity (as we now use the word)?", manuscript, Stanford University.

¹³ Of course, there is variation in the degree to which individuals "take the nation personally." See Waldemar Lilli and Michael Diehl, "Measuring National Identity," Working Paper No. 10, Mannheimer Zentrum für Europäische Sozialforschung, 1999. I would argue that national politicians are especially likely to "take the nation personally," for it is expected of them and besides, their personal fate is largely tied to the fate of the nation.

¹⁴ Fearon, op. cit., p. 11.

¹⁵ Human boundaries must be significant to national self-definition, but territorial boundaries may or may not be. If they are significant, they can be classified under the heading "national interests."

attributes. The equivalent of "principles of action" on the national level might be "causal beliefs" or "world views."¹⁶ I intentionally omit this from the concept in order clearly to differentiate national identity from ideology, strategic culture, operational code and related concepts (discussed below, in the section on "differentiation"). The third change is the explicit inclusion of unconscious levels of identity. Fearon seems not to allow for this possibility with the phrase "a person thinks," but then he actually sneaks it back in with his heading (b). Fourth is the inclusion of "shame," the opposite of "pride." This notion seems to be indicated by Fearon's heading (c), but it is important to be explicit about pairing shame and pride, in order to emphasize the necessary emotional component of the identity concept.

National identity, like a cubist painting, has multiple dimensions and can look different depending on the angle one takes. So for a study of international politics, it is of use to define not only national identity but national *geopolitical* identity: "National geopolitical identity is the set of elements of national identity that cue or are cued by national self-comparison with extra-national human groups and other international structures and processes."¹⁷

Correspondence on the operational level

On the operational level, "correspondence" has to do with the "fit" between the posited variables and the way they are measured. In other words, to borrow statistical terminology, the goal of measurement validity is the goal of minimizing "bias."¹⁸ In the best case, the variable in question would be "directly" observable, but as noted above, in the social sciences this is not very common. So the posited underlying, unobservable variable must be measured by its presumed observable causes or effects.¹⁹ The goal here is to minimize as much as possible the slippage between posited variable and the operational measure. There are at least two ways to do this. First, the *theoretical assumptions behind the measures should mirror the theoretical assumptions about the underlying posited variable*.²⁰ For instance, class can be understood as an identity (e.g., "blue-collar billionaire") or as a simple reflection of material wealth. If we hold the former definition of class conceptually, we can hardly use material wealth to measure it operationally.²¹ At the same time, if class is taken as a multi-dimensional concept including both identity and wealth, it is necessary to find measures corresponding to each of these dimensions. Second, one should *minimize the number of additional theoretical assumptions needed for the "auxiliary measurement theory."* In general, to maximize the chance of measuring what we think we are measuring, we should use measures that can be claimed to be as "close" as possible to the posited unobservable variable, that is, that allow the least scope for intervening variables. The best way to understand this point is to look at a violation of it: one political methodologist's suggestion to use "turnout in elections" as a measure of "sense of citizen or civic duty." The

¹⁶ Keohane and Goldstein 1993

¹⁷ One might wonder about the idea of self-comparison with "other international structures and processes." Should not all comparison be between "likes"? But in fact, self-comparison on the individual as well as the group level is not only relative (with other "likes") but also absolute (with unchanging structures). An "A" can be awarded on the basis of a "grading curve" or on the basis of an absolute standard of "what an A paper looks like."

¹⁸ That is, to make it reflect "on average" the "true" value. By contrast, reliability is about minimizing "inefficiency," that is, making its estimates less uncertain.

¹⁹ Blalock, op. cit., p. 34.

²⁰ This point, and the example that follows, is adapted from R. Pawson, "From indicator selection to theory adjudication," in Hox and de Jong-Gierveld, op. cit., pp. 150-2.

²¹ Of course, class identity can be a correlate of material class, but that is different from considering the two as "parts of" the same concept. See Paul Lazarsfeld and Morris Rosenberg, "Introduction," in Lazarsfeld and Rosenberg, op. cit., p. 15.

number of additional assumptions we need to make to use such a measure is simply heroic: we would have to assume that turnout in elections is dependent *only* on "sense of citizen or civic duty" and not on the weather, the importance or closeness of the electoral contest, and literally dozens of other potential factors.²²

In attempting to operationalize the concept of "national geopolitical identity" (NGI) we quickly realize the assumptions behind the measures used in much research on national identity, which takes identity as a discursive structure, are incompatible with the assumptions behind the measures one needs for NGI defined as a "personal." cognitive-affective complex. Studies of collective identity and "memory" as exemplified in monuments, maps, and museums would clearly be of great value to some projects, but not to this one. Thankfully, given the end explanatory goal of this project, the set of individuals whose national identities we must focus on is rather constrained: national leaders.

How should we study those leaders' national identities? The stricture of adding as few additional assumptions as possible tells us that we should look at their *words* more than their *deeds*. We often say that a man should be judged by his deeds more than by his words. But in politics, things are different. Except for special situations (of which I would argue the decision for against nuclear weapons is one), politicians are not free to "do as they please." Indeed, many decisions "by politicians" may not be anything more than a ratification of the existing correlation of forces in society. This is not necessarily the result of a dereliction of duty; as Max Weber normatively argued, for politicians the "ethic of responsibility" trumps the "ethic of conviction."²³ Instead of looking at deeds, therefore, it seems a better strategy to look at what leaders themselves say about their identities, or at what leaders say in contexts in which they are expected to speak for the nation as a whole. These are the statements which require the fewest additional "auxiliary measurement" assumptions bridging from unobservable variable to observed effect. Such situations are relatively common. We should of course view these statements critically, but if the leaders' explicit intention is to provide us with what we want to know, we should listen.

Major national speeches such as "inaugural addresses" and "state of the union addresses" are an excellent place to start, for a number of reasons:

- They are *speeches to, and on behalf of, the nation as a whole*.²⁴ The breadth of the audience provides a check against politicians' "Zelig-like" tendency to portray themselves in different ways to different audiences, and the ceremonial nature of the occasion provides a check against politicians' tendency to focus on what divides them from their domestic political opponents.
- They are *broad statements of national attainments and objectives*. The breadth of issue coverage allows for a test of the relative importance assigned to different issues and relationships at a particular moment in time. The requirement of breadth may well lead to vagueness on specifics, but NGIs are also vague on specifics.

²² Asher, "The research process," op. cit., p. 13.

²³ Max Weber, "Politics as a vocation."

²⁴ For instance, Ernest R. May concludes from his study of American State of the Union addresses that although "in these messages, presidents plugged their programs. They spoke as party leaders. Often, they spoke lines written by others.... Nonetheless, most presidents tried to make these messages addresses from the throne, appealing to and speaking for the nation as a whole." Ernest R. May, "National Security in American History," in Graham Allison and Gregory F. Treverton, eds., *Rethinking America's Security* (New York: W. W. Norton, 1992), p. 95.

- They are speeches with some *historic significance*. It is commonly accepted that national leaders are worried about their "place in history," so when they feel they are speaking to history, they approach their task with a different mindset than when they are giving a stump speech.
- They are *regular events* in the political calendar, often at the beginning of the political or the calendar year, and therefore the wider political context tends to be less charged by short-term strategic goals and bitter rivalries as is the case, say, for an election-eve address. This regularity also has advantages for defining the "population" from which to "sample."

In short, these various contextual factors combine to make "straight talk" in general the *strategic* thing to do in these speeches.²⁵ Thus such addresses represent the purest possible public expression of a leader's national identity. They therefore fulfill the criteria of "correspondence": keeping the measurement as close as possible to the underlying unobservable variable.²⁶ Such addresses are, of course, biased toward explaining political choices in terms of fundamental national interests and values, when in fact the choices may have been made for entirely different reasons. But this is irrelevant to the measurement of the *content* of leaders' national identities, as opposed to the measurement of the effects of those identities. It is more of a problem that leaders may twist their expressions of their national identity in order to fit their short-term policy goals. But this is more a matter of measurement inefficiency than of measurement validity, assuming that there is any core identity at all to be twisted. Awareness of this potential for "spin" therefore simply militates for the collection of multiple indicators (for instance, looking at more than one speech per politician). This problem is also lessened to the extent that we believe that we can tap an "unconscious" level of national identity which, since it is unconscious, is unlikely to be manipulated.²⁷

This is not to dismiss all criticism of using public speech to gain understanding of politicians' beliefs. To the extent possible, it is also wise to look for "private" expressions of national identity, whether in contemporaneous diaries, correspondence, or through interviews with the politicians' close associates, and then to correlate this data with the data from public speech.²⁸ I will delve further into the notion of multiple measures and methods in the section on "robustness."

²⁵ Pool makes a useful distinction between "instrumental" and "representational" models of communication, but these two models need not always clash.

²⁶ In addition to these advantages, there are advantages of using these speeches on the other two "validity" dimensions, "differentiation" and "robustness." On the level of "differentiation," it is precisely this type of document that "identity as discursive structures" scholars use to do their analysis. This permits a "truce" between "micro" and "macro" conceptions of identity on the issue of the location of data appropriate for theory adjudication. Moreover, an advantage on the level of "robustness" is that most independent nation-states have such speeches as regular parts of their political calendars. This permits for cross-case comparison.

²⁷ By "unconscious identity" I mean not necessarily a psychodynamic idea, but rather a set of statements that are taken by an individual or by a society as simple "facts" but actually reveal certain interpretive choices and emphases.

²⁸ Memoirs may be another measure, though they are still public speech and thus may not "tell all." But a more significant drawback of using memoirs is that people's national identities may evolve over time, and they may then read back into their earlier behavior an identity that they only developed later. I will return to the issue of national identity change below.

IV. Differentiation

Differentiation on the conceptual level

If "correspondence" is having a clear idea of what we want to measure and coming up with the best ways of measuring it, "differentiation" is having a clear idea of what we do *not* want to measure and how to distinguish it operationally from what we do want to measure. On the conceptual level, this means *clearly distinguishing between the definition of the posited unobserved variable and the definitions of other, theoretically similar variables*. For instance, a high level of spirituality is said to be a key component of both "post-materialist" and "traditional" values. In order to differentiate between these two concepts on this dimension, we need to find some way in which they differ in the idea or expression of spirituality. We can also decide there is no difference between them on this conceptual dimension, but we must assert a difference on *some* dimension, or else there is no need for two concepts. These theoretical distinctions should not be too subtle to permit adequate measurement, but on the other hand concepts should not be redefined to such an extent that they become "straw men."

There are a number of popular terms in social scientific discourse that are close enough to "national geopolitical identity" (NGI) as to require conceptual differentiation from it. They are similar in that they all have to do with how the external world is perceived. One set of concepts can be grouped under the heading "foreign policy belief systems," including "operational code," "foreign policy ideology," "world views," and "strategic culture."²⁹ These are conceived as pictures of the world and of national interests that are formulated from first principles in a more or less rational and deductive way. Another set of concepts can be grouped under the heading "discursive identity structures," including the notion of a "state identity" that responds to the norms of "world society," and the notion of a "national identity" that responds to the "collective and institutional memory" of domestic society.³⁰ These are conceived as unwritten but relatively coherent systems of discourse and "praxis" that insinuate themselves into nation-state preference formation across the whole spectrum of issues. Finally, a third set of concepts can be grouped under the heading "independent ideas," including individual "foreign policy ideas," "enemy (and other) images," and "historical analogies."³¹ These are conceived as freely floating ideas that individuals attach almost randomly to particular policies but that may later become "sticky."

²⁹ Alexander George, "The 'Operational Code': A Neglected Approach to the Study of Political Leaders and Decision-Making," *International Studies Quarterly* Vol. 13, No. 2, pp. 190-222; Stephen G. Walker, "A Cautionary Tale: Operational Code Analysis as a Scientific Research Program," manuscript prepared for Colin Elman and Miriam Fendus Elman, *Progress in International Relations Theory*, 1999; Alexander George, "Ideology and International Relations: A Conceptual Analysis," *Jerusalem Journal of International Relations* Vol. 9, No. 1, pp. 1-21; Judith Goldstein and Robert O. Keohane, eds., *Ideas and Foreign Policy: Beliefs, Institutions, and Political Change* (Ithaca: Cornell University Press, 1993); Alastair Iain Johnston, "Thinking about Strategic Culture," *International Security* Vol. 19, No. 4 (Spring 1995).

³⁰ Alexander Wendt, "Anarchy is what states make of it: the social construction of power politics," *International Organization* Vol. 46, No. 2 (1992), pp. 391-425; Peter J. Katzenstein, ed., *The Culture of National Security: Norms and Identity in World Politics* (New York: Columbia University Press, 1996); Rodney Bruce Hall, *National Collective Identity: Social Constructs and International Systems* (New York: Columbia University Press, 1999); Gili S. Drori, "The National Science Agenda as a Ritual of Modern Nation-Statehood: The Consequences of National 'Science for National Development' Projects," unpublished dissertation, Stanford University, 1997; James D. Fearon and David D. Laitin, "Violence and the Social Construction of Ethnic Identity," *International Organization*, Vol. 54, No. 4 (Autumn 2000), pp. 845-877.

³¹ Keohane and Goldstein, *Ideas and Foreign Policy*, op. cit.; Richard K. Herrmann and Michael P. Fischerkeller, "Beyond the Enemy Image and Spiral Model: Cognitive-Strategic Research after the Cold War," *International Organization* Vol. 49, No. 3 (Summer 1995), pp. 415-450; Yuen Foong Khong, *Analogies at War: Korea, Munich, Dien Bien Phu, and the Vietnam Decisions of 1965* (Princeton, NJ: Princeton University Press,

"National geopolitical identity" as conceived here is different from all of these other concepts. The differences can be shown most clearly in a 2 x 2 table:

Table 1: Conceptual Distinctions Between Constructs About Perceptual Frames

	Top-Down Formulation	Bottom-Up Formulation
Identities	Discursive Identity Structures	<i>National Geopolitical Identity</i>
Ideas	Foreign Policy Belief Systems	Independent Ideas

National geopolitical identity differs from both discursive identity structures and foreign policy belief systems in that it is "bottom-up," in two senses. First, *NGI is bottom-up in the sense that it is located at the micro-level rather than the macro-level*. This does not mean that all individuals' national identities differ, nor does it mean that individuals fail to make use of the cultural materials around them in society. What it does mean is that identities should be measured one individual at a time, and that identities are matters of true personal self-expression rather than mere roles or stances that individuals take on in certain social contexts.³² Therefore, in order to distinguish NGI from these other variables, the bottom line on this dimension is to look for consistency of perceptions by an individual across different contexts (e.g., public and private) and across time, as well as for differences between the perceptions of at least some individuals at any one time. Second, *NGI is bottom-up in the sense that it need not be coherent across external objects*. The phrase "international politics is marked by zero-sum conflict" is an expression of a foreign policy ideology which is built from first principles, not of an NGI which builds different images of different relationships through inductive experience. Therefore, in order to distinguish NGI from these other variables, the bottom line on this dimension is to look at whether all external objects are viewed through the same perceptual prism, for instance whether the perception of the level of conflict depends much more on the perceived nature of the issue-area than on the perceived historical behavior of the opponent.

While NGI differs from discursive identity structures and foreign policy belief systems because of its "bottom-up" nature, NGI also differs from both independent ideas and foreign policy belief systems because it is an "identity." There are at least two main differences between NGIs and these other "ideational" constructs. The first difference is that national geopolitical identities are about *self-definition*, whereas foreign policy ideas are "theories" that can be tested against experience. The Munich analogy in the Vietnam case was in a sense a "theory" about how the war might evolve, a theory that could be discarded upon the accumulation of disconfirming evidence. There are many reasons why such ideas are not discarded soon enough, but as they are theories they are much more easily discarded than elements of self-definition (for which there is essentially no empirical test). Therefore, in order to distinguish NGI from these other variables, the bottom line on this dimension is to measure whether or not the speaker displays not only inconsistent portrayals of different external objects, but also consistent portrayals of the same external objects across issue-areas and across time.³³ The second point

1992); Richard E. Neustadt and Ernest R. May: *Thinking in Time: The Uses of History for Decision-Makers* (New York: The Free Press, 1986).

³² Paradoxically, the posited existence of unchanging discursive identity structures has been seen as allowing a great degree of rational "picking and choosing" of elements of those structures in order to further short-term political aims. With such a conception, one wonders, along with Fearon and Laitin, what outcomes the identity variable could possibly explain. Fearon and Laitin, "Violence and the Social Construction of Ethnic Identity," op. cit.

³³ There must be of course some potential in the theory for identity change. This is a big subject and one that I cannot tackle here in any depth. As a first cut, I would suggest the hypothesis that identity, like "culture,"

that differentiates NGIs from these other ideas is that identities are not merely cognitive but are also *emotional*.³⁴ This has significant behavioral consequences, but it also has significant consequences on the level of expression. The "model of man" suggested by the "ideas" approaches is of a rational actor who takes cognitive shortcuts. Such an actor has no need to use emotion-laden vocabulary in discussing national options and choices, has no need to tell a moving story about a letter Washington sent from Valley Forge, and has no need to stand with his hand over his heart as the national anthem is sung. Therefore, in order to distinguish NGI from these other variables, the bottom line on this dimension is to look for emotive language, for references to national symbols, and for references to historical or mythic events in the past such as the founding of the nation.

Differentiation on the operational level

It is also important to have differentiation on the operational level between unobservable variables that are theoretically similar. I have made a number of points in the preceding subsection on what to look for to establish differentiation; now the question becomes with what instruments to do the looking. The best way to do this is to make sure to develop measures of what theoretically distinguishes these variables using the *same measuring instruments on the same data set*. The smaller the difference in measurement instruments, the clearer the (potential) difference in the measurements themselves—all the way to "Variable A is here at this level, but I find no evidence of the presence of variable B."³⁵ Borrowing the data sets, methods and instruments put forward by "proponents" of other variables is a particularly good strategy, because it neutralizes arguments that theoretical expectations influenced the coding.³⁶ However it is not right to take this strategy to mean simply relabeling a measure meant to reflect one concept as reflecting a different concept, for the goal of differentiation demands that each concept be operationally distinct.

The concept of identity advanced here is broadly compatible on the operational level with the *quantitative content analysis methods* that scholars have employed in the past to measure individuals' "operational codes" and "belief systems."³⁷ People express their identities through words. Quantitative content analysis can serve systematically to transform these words into transparently constructed and replicable "data sets" that are susceptible to various hypothesis tests and tests of statistical significance. The basic method of quantitative content analysis is to break down a text or set of texts into smaller units (words, phrases, or in my case paragraphs) and to measure the relative frequencies of different objects of interest across those units. The method can sometimes be criticized for laboriously proving the obvious, but that critique can be extended to much of social science. Moreover, the critique is somewhat unfounded, for it is not the case that looking at the "surface" of a text must only find what the speaker or author wanted

changes only in the face of major personal or national crisis.

³⁴ The emotional level of identity—in particular the need for self-esteem—also adds to the force with which people cling to key elements of self-definition.

³⁵ On a more philosophical note, holding the instruments constant is useful because it allows for theory adjudication even in the absence of the potential for theory confirmation or falsification. The point is rather Habermasian: in the absence of "hard" evidence, scientific progress can still take place if proponents of different theories can agree in advance on what will constitute evidence. Pawson, "From indicator selection," *op. cit.*, p. 145.

³⁶ King, Keohane, and Verba, *Designing Social Inquiry*, *op. cit.*, p. 157.

³⁷ As for instance in the classic study of John Foster Dulles by Ole Holsti, "Cognitive Dynamics and Images of the Enemy," in D. Finlay, O. Holsti, and R. Fagen, *Enemies in Politics* (Chicago: Rand-McNally, 1967), pp. 25-96.

us to see. In fact, content analysis categories, if carefully constructed, can tell us more about the speaker than the speaker actually understands about himself. For instance, basic issues of self-definition and self-esteem may simply seem to be "facts" to individuals rather than matters of interpretation and choice. Therefore they may in fact not be aware of how revealing these "facts" are about their true identities, and not being aware of "showing their hand," they will not try to cover it up. Therefore, in order to understand how individuals see themselves and the world, it is often better to look at what they *say* and not just at what they *mean*.

"Identity" scholars *per se* have tended to ignore the benefits that quantitative content analysis techniques can offer.³⁸ This is partially due to issues related to the sociology of the social sciences, and partially due to a real conceptual divide. The macro-sociological understanding of identity emphasizes deep narrative structures that are not immediately apparent, while quantitative content analysis almost invariably relies on small, "surface-level" units of analysis (the word, sentence, or paragraph). The search for deep narrative structures is neither compatible with the fundamental need of quantitative analysis for "large n," nor with the fundamental assumption of quantitative content analysis that the *frequency* of occurrence of given "surface-level" content characteristics is important.³⁹

But on the other hand, the great advantage of quantitative content analysis is that it guards against ill-conceived interpretations. "Nonfrequency" content analysis is often based on a small or quirky "data set" which offers few checks against the analyst's seeing what he wants to see. Moreover, given its goal of diving "beneath" the text, in practice it gives the analyst almost unlimited license in fashioning interpretations. Any utterance, however rare, and even any silence can be taken as tremendously significant indicators of "deep" identity. It goes without saying that some of these interpretations can seem more literary than social-scientific. In addition, even if the "nonfrequency" method is employed with care, by documenting and ranking all the plausible interpretations that can be found in the text, it still faces an uphill battle to convince the skeptical reader of the correctness of these judgments, for the "decision rules" favoring one interpretation over another are rarely transparent. The very method of quantitative content analysis forces the analyst to avoid these errors.

This is not to say that the qualitative approach is irrelevant to the measurement of identity. As stated above, it is important for operational "differentiation" that "competing" variables such as the notion of "discursive identity structures" be measured and compared with one's "own" variable, if possible on their own "turf." If only for this reason, the qualitative nature of much research on identity would demand that I too take a qualitative look at the data.⁴⁰ But there are other, even stronger reasons for an integrative quantitative and qualitative approach to identity measurement, covered in the section on "robustness."

Having explained the merits of the quantitative content analysis approach, I now proceed to a description of the principal measures taken in this research on NGI. There are five such measures:

³⁸ Some recent examples of a return to quantitative content analysis can be found in Rawi Abdelal, Yoshiko Margaret Herrera, Alastair Iain Johnston, and Terry Martin, "Treating Identity as a Variable: Developing Technologies for Measuring the Content, Intensity, and Contestation of Identity: A Proposal," unpublished manuscript, Harvard University, October 2000.

³⁹ See George, Alexander, "Quantitative and Qualitative Approaches to Content Analysis," in Ithiel de Sola Pool, ed., *Trends in Content Analysis* (Urbana: University of Illinois Press, 1959), p. 7-32.

⁴⁰ At the same time, "macro-level" identity concepts should also try to establish themselves against "micro-level" identity and other ideational concepts, so they should engage in the same multi-method approach as that suggested here.

- A. Who is out there (outside our boundaries);
- B. Who is "us," what are our proper "human" boundaries;
- C. What issues we consider to matter most;
- D. Interest and values distance between "us" and "them";
- E. "Our" potential impact on "them", "their" potential impact on "us."

To understand why these measures have been chosen, recall the definition of national identity: "national identity is the set of people, values, interests, and capabilities by which a person consciously or unconsciously defines her national group and about which she feels some level of pride or shame," and the definition of national geopolitical identity: "national geopolitical identity is the set of elements of national identity that cue or are cued by national self-comparison with extra-national human groups and other international structures and processes." It should be immediately apparent from these definitions why most of these measures are necessary. Measure A has to do with self-comparison to external groups. If we are not considering the relationship of our nation to external groups or structures, then the statement should not be considered to emanate from the speaker's national geopolitical identity. Measure B has to do with human boundaries, the set of people by which the individual defines her group. Measure D has to do with values and interests of the nation. While it would be of interest to distinguish between these, in practice I have found it essentially impossible to do so. Measure E has to do with the capabilities of the nation.⁴¹ Measure C is the "oddball" here, included more to permit differentiation from other similar concepts, rather than as an integral part of NGI itself. The precise coding rules for these measures are in the appendix to this paper.

What dimensions of national geopolitical identity do these measures leave out? Identity, as I have argued, is a cognitive-*affective* complex in which emotions are not only present but can even take center stage. Yet it is quite difficult to devise rigorous and systematic quantitative techniques to unearth this emotional side. Looking for emotional keywords might offer some clues, but such a procedure can lead to misleading results. We often think of emotions as something people wear on their sleeve, but in fact the presence of emotions can be much more subtle. People cannot learn to be unemotional, but they can learn to hide or to flaunt their emotions.⁴² These are matters for continuing study and methodological creativity. For now the best I can do is to offer frequency of mentions of an issue or of an external "Other" as an imperfect proxy for degree of emotional investment in it.⁴³

Other challenges which I have not fully met are the following: (a) the question of creating indices out of these various measures—for instance, is a medium-focus/high-conflict relationship more or less consequential for self-definition than a high-focus/medium-conflict relationship?; (b) the question of establishing qualitative thresholds within the quantitative results—at what point does an identity become "oppositional nationalist?"; (c) the question of proper statistical

⁴¹ This is a concept developed by Albert Bandura. See Bandura, Albert, "Exercise of Personal Agency Through the Self-Efficacy Mechanism," in Ralf Schwarzer, ed., *Self-Efficacy: Thought Control of Action* (Washington, DC: Hemisphere Publishing Corp., 1992).

⁴² I would argue that the degree of intensity with which a politician makes any *single* statement is more revealing about his strategy than about his underlying beliefs. We can think of the politician as a c.d.: he will generally say the same things, but he can say them at different volumes. For a review of the psychological literature see Johnstone, Tom and Scherer, Klaus R., "Vocal Communication of Emotion," in Michael Lewis and Jeannette M. Haviland-Jones, eds., *Handbook of Emotions*, 2nd ed. (New York: The Guilford Press, 2000), pp. 221-235.

⁴³ This idea is suggested by Ithiel de Sola Pool, "Trends in Content Analysis Today: A Summary," in Ithiel de Sola Pool, ed., *Trends in Content Analysis* (Urbana, Illinois: University of Illinois, 1959), pp. 194-195.

significance tests to establish confidence scores for various findings. It is my hope that with more experience with the technique, some of the answers to these questions will "organically" appear.

V. Robustness

Robustness on the conceptual level

The third key aspect of validity is "robustness" or the degree of generalizability. On the conceptual level, this means at a minimum *defining the unobserved variable in generic terms*, that is avoiding defining it in terms of proper nouns and thereby permitting multiple observations.⁴⁴ Failure on this score means a failure to create a variable at all. Beyond that minimum standard, establishing robustness on the conceptual level means *making the case for the potential presence of the variable in a clearly defined set of contexts (or "domain")*.⁴⁵ In order to establish robustness it is not necessary to establish that the variable is everywhere present, but the variable would seem more important to the extent that it is not too narrowly contextual.

Having already defined NGI generically, I can immediately plunge into the establishment of its robustness by defining the domain in which it may exist. My conception of identity is based on seemingly basic elements of human psychology: the need for "collective self" and other-categorization, self- and other-definition, and self and other-esteem.⁴⁶ The precise way in which these basic needs take form may depend on the individual or the historical era in question, but the needs are part of human nature. Aristotle took as the fundamental assumption of his political science that "man is a political animal" who can only realize himself through the pursuit of the common aims of his *polis* such as security and welfare. Even Hobbes recognized that the notion of the atomized "state of nature" that preceded the creation of human society is an unhistorical hypothetical. At the same time, the idea of a *national* identity is clearly a historically bounded concept. As many historians have pointed out, the "nation" as a sovereign political community with relatively fixed human boundaries took root only in the 18th and 19th centuries, first in Europe and then gradually throughout the globe.⁴⁷ It would be invalid, therefore, to try to analyze the "NGI" of a French nobleman of the 15th century, or even of a French peasant of the mid-19th century.⁴⁸

The fact that an individual has an NGI does not mean that it is her only collective identity. National identity can co-exist with both sub- and supra-national collective identities, and it can exist whether or not the individual feels the nation or the "nation-state" to deserve her "supreme loyalty."⁴⁹ In some contexts, the individual may feel a tug of loyalties to different levels; in other contexts, the individual may feel the loyalties to be mutually reinforcing; in still other contexts, the individual may perceive the situation from only one "level" of identity. The important thing is to demonstrate the validity of the nation as *a* level of identity that is cued by

⁴⁴ On this point, see Adam Przeworski and Henry Teune, *The Logic of Comparative Inquiry* (New York: Wiley-Interscience, 1970).

⁴⁵ Brinberg and McGrath, "A Network of Validity Concepts," op. cit., p. 18.

⁴⁶ The psychological ideas underlying this section are associated with Tajfel and Turner's "social identity theory" and "self-categorization theory."

⁴⁷ As should be clear, I borrow my definition of "nation" from Benedict Anderson, *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (London: Verso, 1991).

⁴⁸ The classic work on the "domestic" expansion of the national idea in France is Eugen Weber, *Peasants into Frenchmen: the Modernization of Rural France, 1870-1914* (Stanford, Calif.: Stanford University Press, 1976).

⁴⁹ Hans Kohn, *Nationalism: Its Meaning and History* (Princeton, N. J. : Van Nostrand, 1955), p. 9.

the context in question, not as the *only* or the *primary* level of identity cued by that context. So a statement by Tony Blair about "the British national interest" can be taken as a statement of national identity, even if it is also a statement of Blair's identity as a Labour Party member and as a European. But a statement by King Fahd about "the Saudi national interest" can be doubted to represent a statement of national identity, given that the House of Sa'ud is a family, not a nation.

Robustness on the operational level

On the operational level, establishing "robustness" implies a number of steps. First, the posited unobservable variable should lend itself to the production along each of its key dimensions of *multiple indicators (or "observations" ⁵⁰) for each descriptive inference to be made.⁵¹ Second, it is also of interest to go beyond multiple indicators to using *multiple measures and even multiple types of research methods* to estimate the "true" value of the posited unobserved variable.⁵² Indeed, the ability of a posited unobservable variable to produce multiple, convergent indicators based on multiple measures and methods is the very essence of "robustness."⁵³ To maximize measurement validity is to minimize bias, and since different measures and research methods have different inherent biases, it is therefore necessary to balance them against each other through a process of triangulation. A third element of "robustness" on the operational level is *the production of indicators in varied, accessible contexts.*⁵⁴ Indeed, establishing the validity of the variable in certain specified contexts, taken alone or in tandem, might be enough to allow for extrapolation to a well-defined set of other contexts.⁵⁵*

I employ numerous measurement strategies to establish the plausibility of the claim for robustness of the NGI concept. First and most simply, I attempt to find several "statements of national purpose" from each of the individuals I am analyzing, and my inference about their NGIs is the result of an "average" of these multiple indicators.⁵⁶ Second, as previously mentioned, I attempt to balance my quantitative findings with a qualitative analysis of the same

⁵⁰ Here I am following the terminology of Lazarsfeld and Rosenberg, "Introduction," op. cit., p. 16.

⁵¹ The production of multiple indicators is of course important to demonstrate measurement reliability, but here I am making a point about validity: if the operationalization of the construct is so narrow as to make it synonymous with a single indicator, then it is not proper to speak of it as a variable. Moreover, in the wider sense of validity of causal claims, it can be shown that inefficiency in the measurement of the independent variable can create downward bias in the estimation of the relationship with the dependent variable. King, Keohane, and Verba, *Designing Social Inquiry*, op. cit., p. 158.

⁵² John Brewer and Albert Hunter, *Multimethod Research: A Synthesis of Styles* (Newbury Park, Calif.: Sage Publications, 1989).

⁵³ Donald W. Fiske, "Convergent-Discriminant Validation in Measurements and Research Strategies," in Brinberg and Kidder, eds., *New Directions for Methodology*, op. cit., esp. p. 89.

⁵⁴ Accessibility of context is a crucial practical element for establishing the plausibility of validity claims. For example, when in 1546 John Heywood wrote, "The moon is made of a greene cheese," the existence of the posited variable "a greene cheese" could not be validated because the only context in which this cheese was thought to exist was in a place to which no one could travel. The story of this saying can be found on [<http://www.straightdope.com/columns/990723.html>].

⁵⁵ Taking a different position, Adcock and Collier argue, "Validity needs to be established one context at a time" (Adcock and Collier, "Connecting Ideas with Facts," op. cit., p. 5). But in fact, I would argue that validation in a carefully selected set of contexts can permit for extrapolation of the validity of the variable to other contexts in which it has not been explicitly "established." At least it should shift the burden of proof to those who believe that the validity of the variable does not extend to one of those contexts. For instance, establishing the validity of a concept of "regime" that permits for three sub-types, democracy, oligarchy, and dictatorship, for Europe in 1900 and in 2000 gives strong plausibility to the claim for the validity of this concept for 1950.

⁵⁶ This idea of "average" can be taken literally or not, depending on qualitative judgments.

texts. The multi-method strategy is an excellent way both of establishing the "reliability" of certain interpretations to the extent that different methods arrive at the same conclusion— a topic I turn to in section VI— and of establishing the robustness of the variable overall by showing how different methods can "find" it and help it to take shape. In terms of validity of NGI, there are indeed issues of narrative structure and of interpretive nuance that a quantitative content analysis is not likely to handle as well as a qualitative approach. To take a simple example, one can imagine a speech by Henry Kissinger starting off with a short statement that NATO was necessary because the USSR was the main antagonist of the US, and then proceeding to devote his attention to concerns entirely internal to NATO. If we used this speech as an indicator of national identity, the quantitative categories as currently structured might be taken to imply that the US saw its main antagonist as France. This obviously faulty interpretation could be easily corrected by a qualitative look at the same text.⁵⁷

The third goal for establishing the plausibility of claims of robustness is to show that the variable can be successfully measured in different contexts, both domestically and internationally. The former goal is probably more important at this stage in the development of the research agenda. I have already mentioned the problems of relying exclusively on public speech for the measurement of a "personal" national identity. I have tried to minimize these problems by selecting carefully the public speeches to be analyzed, and by subjecting this data to critical scrutiny. But the fact remains that politicians are intelligent dissemblers and can fool even the careful analyst, appearing sincere while they mouth words that are purely crafted for political advantage or that merely represent inherited patterns of discourse that are empty in content.⁵⁸ It is therefore necessary to look for some evidence that the NGI expressed in public was actually representative of the politician's "true" beliefs. This can be done in some cases through looking at private papers from the period and interviewing close associates, as well through reading memoirs and the secondary literature. It is generally much harder to quantify the results of this work, as careful sampling is difficult when the overall domain of potential sources is so diffuse and perhaps mostly inaccessible. Therefore, such "detective work" should not be taken as a substitute for the analysis of public speech, and indeed the various scraps that it might turn up are likely to provide (highly inefficient) evidence in favor of any number of different interpretations of NGI. Nevertheless, such a procedure can act as a check against a too-naïve view of public speech, and the convergence of measures of NGI from "public" and "private" sources would allow for the generalization of the NGI interpretation beyond the initial context of major public occasions.⁵⁹

As for the latter goal of geographical generality, in my work I am measuring NGI for the cases of the Prime Ministers or Presidents of Australia, Argentina, France, and India. At this stage, it is necessary to establish the concept "one context at a time," so I do the work on each country separately and then look for confirmation of the validity of the overall exercise by sending it to country experts.⁶⁰ However, at some point in the process this practice will hit the point of diminishing marginal returns. It may still be necessary to establish that NGI is a

⁵⁷ This example does not, however, invalidate the *careful* use of a quantitative analysis. It would be interesting to show that the US considered France as its main antagonist within the Western bloc.

⁵⁸ One thinks for instance of Catholic parish priests and parishioners who hardly understood the Latin prayers they once had to utter at Mass.

⁵⁹ This would also help to distinguish between identity as discursive structures and identity as personal cognition and affect.

⁶⁰ I also ask them for their impression of the correctness of the particular interpretation, but this is a question of reliability more than validity.

worthwhile concept in other world regions, such as Africa, but it seems unnecessary to devote as much time to this in future extensions of the procedure to the more established nation-states.

VI. Reliability

Reliability is achieved when "different measurements of the same phenomenon yield the same results."⁶¹ While establishing validity is important at both the conceptual and operational levels and must always rely to some extent on rhetorical persuasion, establishing reliability is important at the operational level alone and can be achieved (especially in terms of quantitative data) via a more mechanical process. I have already briefly touched on how the convergence of interpretations from different methods can indicate measurement reliability. However, I prefer not to use this as the main indicator of reliability, because the main point in using different methods is to improve validity by balancing *divergent* results through an engaged critical eye.⁶² I therefore focus my discussion of ensuring reliability to procedures that are internal to the methods I employ. There are two primary such procedures: (a) increasing the "n" of observations to minimize the effects of random measurement error induced by "nature"; (b) performing an "inter-coder reliability test" to estimate and then minimizing the effects of random measurement error induced by vague or hard to implement coding rules.

There are three levels at which we can think of "raising the n" in the measurement of an individual's NGI: within a speech, across speeches, and across individuals. First, "raising the n" within a speech means coding smaller units rather than larger ones. In my quantitative analysis I have done this by coding speeches by paragraph instead of by larger units. At the same time, one should not place the goal of reliability over the goal of validity. I have established as my basic unit of measurement the paragraph rather than the individual word, because it is appropriate to take ideas as the "atoms" of NGIs, and paragraphs are generally the grammatical unit that best corresponds to a single idea. Second, "raising the n" across speeches means coding multiple speeches by the same individual. I have tried to do this as much as possible. I have found that each "state of the union" address requires between 1-3 hours of coding and sometimes an equivalent amount of time simply to be located, so a dramatic increase in the "n" at this level is not possible without additional manpower. In addition, the issue of validity raises its head here too; it would not be wise to try to "raise the n" by extending the quantitative analysis to narrow policy speeches made to narrow audiences, as this practice would introduce bias and even inefficiency. Third, "raising the n" across individuals might not seem necessary for coding the NGI of one individual, but in fact it is a good idea, for we are often interested not only in understanding the NGI of one individual in isolation, but also in understanding how much that NGI differed from his predecessors in power or other major political figures of the day. It is therefore necessary not only to "raise the n" for him but also for these others around him. Of course, issues of validity also apply to this practice.

The other major avenue for measuring and improving reliability is to perform an inter-coder reliability test. The inter-coder reliability test is a test of the extent to which coding decisions can be replicated by different coders.⁶³ This is a crucial step for establishing that the

⁶¹ King, Keohane, and Verba, *Designing Social Inquiry*, op. cit., p. 151.

⁶² King, Keohane, and Verba recognize the potential contradiction of the goals of validity and reliability. They seem to imply on p. 151 that one may "purchase" reliability "at the expense of validity," but this goes against their overall argument that bias is more problematic than inefficiency. Ibid., p. 151.

⁶³ Stephen Lacy and Daniel Riffe, "Sampling Error and Selecting Intercoder Reliability Samples for Nominal Content Categories," *Journalism and Mass Communication Quarterly*, Vol. 73, No. 4 (Winter 1996), pp. 963-973.

coding rules as defined are transparent and straightforward to implement. A major problem in both quantitative and qualitative analysis of texts is that the results of the coding may depend on the personality of the coder. The inter-coder reliability test helps to measure the extent to which this may be true. The general practice in such tests is as follows. At least two coders, generally versed in the subject matter, should be trained in the coding rules and then asked to code a randomly selected sample of the data. The proper size of the sample can be determined according to the overall size of the population being analyzed, the degree of acceptable probability for estimating the confidence interval (often 95%), and the minimum acceptable correlation between codings (often 80%).⁶⁴ The correlation in the codings is then computed. In fact, roughly 85% of codings should be found to be in agreement in order to have 95% confidence that the score of 80% or above was not the result of random error.⁶⁵

The inter-coder reliability test has clear limitations for establishing the "value" of the coding scheme. The results of a coding scheme may not depend on the personality of the coder, but they may still depend heavily on the personality of the author of the rules. This brings us back to issues of validity. One mechanical way to test the "legitimacy" of the results—uncommon in the literature but probably fruitful—would be to perform a "sensitivity" analysis to test how important various coding rules are for the overall results obtained. If one knew in advance which coding rules were likely to raise the most objections, one could offer a measure of the extent to which the more "questionable" coding rules drive the results.⁶⁶

[What follows is a description of the actual practice and results of the inter-coder reliability test]

⁶⁴ Ibid., p. 967.

⁶⁵ Ibid., p. 966.

⁶⁶ In terms of my own coding scheme, often several rules can be said to point to the same coding result. I am sure that it would be quite tiresome to write out for each individual coding decision the full list of all the rules that point in a certain direction. It might nevertheless be interesting to experiment with this.

APPENDIX: CODING RULES FOR IDENTITY

Detailed description of each question

The questions of measures A and C are easily understood. They are, "Who/what outside our borders do we refer to, and in what issue context?" Answering these questions simply requires us to count up the references to external actors and to issue-areas. I take the paragraph as my unit of analysis, so multiple references to the same actor or issue-area in the same paragraph would count only once. However references to different actors/issue-areas in the same paragraph would each count in the coding scheme. The external actors could be other nations, other non-national groups such as multinational corporations, collections of nations such as NATO or the "First World," or impersonal international structures such as the world economy. The issue-areas could be economic, domestic non-economic, or international non-economic. For example, the phrase of Argentine President Fernando De La Rúa's March 2000 state of the union address, "We will go out with Mercosur to capture new markets," would be coded as an outward-looking reference both to the Southern Cone⁶⁷ and to the wider world ("new markets"), and as being in the economic issue-area.

The question of measure B is the question of our "human" boundaries. In the NGI context, the question arises mainly in terms of whether or not "others" are really not "others" at all. The precise question I ask is, "Are the nation and the "other" seen as a unitary actor, or at least as linked by bonds of family?" This is a restrictive coding scheme that does not code mere assertions of common interest as evidence of a wider identity. So in the De La Rúa quotation above, "We will go out with Mercosur to capture new markets," the identity expressed is simply Argentine, because he differentiates between Argentina and Mercosur and because there is no emotional content to the Mercosur attachment.

The question of measure D is the question of self-definition beyond the simple establishment of boundaries. We often want to know the content of interests and values, but in the NGI/foreign policy context the key question is rather how "far" our interests and values are from those of others, more than the specific content. The perceived "distance" between us and "others" in terms of interests and values is a good proxy for friendship and enmity. We can pose this question as, "How good is the (potential) fit between our values and interests and our environment (understood as a whole, as parts of the whole, or as a particular part of the whole)?" In theory it would be better to distinguish between "interests" and "values," but in practice, in coding this distinction is quite impossible to make. The coding for each mention of the outside world is a simple dichotomous one: "conflict" (a score of 1) or "no conflict" (a score of 0). In a few ambiguous cases there could be a coding of "mixed conflict/no conflict" (0.5) but this coding should be used as sparingly as possible. Once again returning to the quotation from De La Rúa, "We will go out with Mercosur to capture new markets," I code this as "no conflict" vis-à-vis Mercosur (for obvious reasons) and "conflict" vis-à-vis the wider world (because the word "capture" is aggressive and implies a zero-sum contest with other economic competitors).

The question of measure E is the question of self-efficacy: national strength or weakness. For the NGI case, the question of self-esteem can be posed as, "To what extent can we (potentially) impact our environment, to what extent can we and our environment mutually impact each other, and to what extent can our environment impact us?" By "impacting" I mean having an effect on the other outside of our own borders. This effect could be material—we win

⁶⁷ Certainly, not all Southern Cone states are full members of Mercosur. But it seemed more useful to track references to the group of states in the Southern Cone over time, including before there was a Mercosur.

a contract they sought—or ideal—they learn how to run a democracy from watching us. The coding for each mention of the outside world would be "we impact" (a score of 1), "they impact" (a score of -1), or "mutual impact" (a score of 0). Note that in the case of positing some other group as structurally in the same position as ourselves, I code this as "mutual impact." For we have a kind of relationship of equality even though we are not seen to be in direct contact with each other. So again to return to the De La Rúa example, "We will go out with Mercosur to conquer new markets," there is a coding of "mutual impact" vis-à-vis Mercosur, because there is a relation of equality and mutual help; and there is a coding of "we impact" vis-à-vis the wider world, because he is claiming that as a result of Argentine action, the world economic landscape will change.

Precise Coding Rules as of 4/16/01

I. General rules for all codings

0. The basic unit of measurement is the paragraph, under the assumption that the format follows the dictionary definition of a paragraph as "a distinct division of written or printed matter that begins on a new, usually indented line, consists of one or more sentences, and typically deals with a single thought or topic." If this assumption is routinely violated, then a different unit of measurement should be established to keep each unit related to a single thought or topic.

1. Statements of how things were in the past, how they are now, and how they will be in the future are all coded alike.

2. However, statements of secular trends should be taken into consideration in the coding, with the projected evolution taking precedence over the past or present reality.

3. When in doubt on coding, always stick to the explicit language of the text rather than pouring in dubious "spin" interpretations.

4. Each measure should be coded separately from the others and should not be confused with the others. The measures are designed to be at least potentially uncorrelated.

5. Coding choices need not be consistent across paragraphs or other content units. Each unit should be viewed as a largely self-contained unit. Information from other paragraphs should not be brought to bear on the coding of the paragraph at hand, unless the speaker is clearly referring back to other paragraphs, or there is simply not enough information to code successfully without reference to previous paragraphs.

II. Coding for Outward-looking references

1. Count any explicit reference to anything explicitly foreign, whether it's seen as inside or outside our national boundaries. Don't count any reference to anything that isn't foreign, e.g. to members of the national group who happen to be found outside the national territory (for instance, references to 'smugglers' may not count, if the nationality of the smugglers is not clear).

2. Count any explicit reference to a world beyond our boundaries, be it another country or countries, a geographical unit of which we're a part, an international structure/movement/trend/development/wave, or a general case of a generic community/country/nation in the world (but note on the last point: do not count references that could be understood as referencing a philosophical ideal without a clear international

perspective—so, in practice, references to “a nation/country” would generally count, while references to “a society/an economy/a democracy” would generally not count).

3. Implicit references to others, to count, must be very clear and unambiguous. Usually this implicitness will be due to the fact that the explicit reference was just made in the preceding paragraph and thus it is not necessary to repeat it.

4. Even if the external orientation of the implicit reference is clear, if the precise external object referenced is not clear then it shouldn't be counted. For instance, calls for national “liberation” may not be clearly externally directed—but even if we decide that a certain call is externally directed, if the external or foreign obstacle to liberation is not easily determined (another country? A multinational corporation? The world economy?) then the reference shouldn't be counted. It should however be tallied to see how many references are going uncounted.

5. Don't count any references for which results on the other measures can't be gotten. This often is the case with bland reference to exports/imports statistics. But tally them to see if this is a major problem.

6. The objects of references should be reported exactly as they appear in the text. At a second stage, these references can be grouped according to the speaker's categories (first priority) or other logical categories developed by the author. The source and coding rules of the categories should be clearly stated.

7. The maximum “number” of references to the same object from the same paragraph or other unit of content chosen that can be counted is 1. If the same object is referenced more than once, then the coder will simply have to balance multiple pieces of information in his codings of other measures.

8. There is no maximum number of references to different objects that can come out of the same paragraph or other unit of content chosen.

9. If the object referenced could be either a country/external group, or an impersonal movement/trend/situation, code the reference as to the country/group and not to the impersonal movement/trend/situation.

III. Coding for Interest/Values Distance (i.e., “with” or “against” us)

For a coding of “with”:

0. If the nation finds itself in (or recalls) active cooperation with the other (or concretely proposes such cooperation without setting prior conditions for such cooperation). This coding is for when the cooperation, however difficult or dangerous, is desired by us; it would count as “against” if the cooperation is forced (by us on them or by them on us, e.g., as a result of our military failure) or otherwise is not a reflection of the speaker's idea of the true national interest (e.g. if he is opposing an instance of cooperation supported by others in our nation).

0.5 See “against” coding 0.

1. If our nation is sharing with the outside object one or more of the following: (a) membership in a club, group, movement, trend, wave (even if we criticize that group for some shortcoming); (b) a posited similar structural position and hence common interests (including cases where we explicitly point out such basic common interests in spite of some ongoing conflict. Note, there must be a claim of a basic harmony of interests and potential for cooperation that goes beyond a simple desire for a “halt to the shooting”); (c) a posited common lineage, heritage, or values with the outside object; (d) generic friendship, trust, or merely correct relations.

2. If our nation is engaging the outside object in equal exchanges bringing mutual benefit (a liberal model of exchange).
3. If our nation is facing external penetration/pressure/ignoring/interest/influence/inflexible structures without objection, resistance or complaint. Or is penetrating, pressuring, etc. others without their complaining.
4. If our nation is seen as (a) a model for others or (b) seen as justly taken by itself and others as an anti-model, if (c) others (including a generic other) are a model for our nation; if (d) nation is seen as a leader for others; (e) takes others as a leader.
5. If the other is unilaterally creating positive benefits or opportunities for us even if we haven't yet gotten them in hand.
6. If lack of cooperation with the other is due to failings on our part that we recognize to be failings.

For a coding of “against”:

0. If the nation finds itself in (or proposes getting into, or recalls being in) an active conflict (including refusal to cooperate) with the other. However code as "with" if it is explicitly argued that the conflict is somehow a mistake or otherwise fails to reflect our underlying harmony of interests and values.

0.5. See "with" coding 0.

1. If our nation is not sharing with the outside object any of the attributes mentioned in (1) above.
2. If our nation is engaging in unequal exchanges, the distribution of whose benefits depends on relative power (whether we gain more or lose more) (mercantilist model of exchange). Included in this would be talking about relationship in pure power terms, concern for power balancing, power projection etc.
3. If our nation is facing external penetration, etc. (this includes attempts to separate it from its allies) and objects, resists or complains/ is penetrating others and recognizes that they resist or complain.
4. If our nation is unjustly taken by others as an anti-model (not just ignored); if we take others as an anti-model.
5. If our nation is in a competitive race with others (especially in terms of relative numerical rankings whether we are above or below the others) and/or holds mercantilist ideas (including ideas of "national competitiveness").
6. If the nation issues threats to the other or perceives the other to issue threats to it.
7. If the other is causing us difficulties, costs, problems, or harm even if we can overcome them, and even if they are merely side effects of his policies.

Note that codings 1-4 are mirror images of the “with” codings. Coding 5 is a mere emphasis of (1a): if we are in a competitive race, then we are not members of the same club—my coming in 1st means that the best you can do is 2nd, and vice versa.

III. Coding for impact of us on environment, environment on us

[Note: if we are discussing a past high position from which we've fallen, code high; if it is a past low position from which we've risen, also code high. In general code high. This derives from basic self-esteem theory]

For a coding of “we impact our environment”:

0. If we are clearly monopolizing the role of protagonist in the relationship.

1. If our nation (or some representative of it) makes or has made a contribution to any aspect of life in our environment (history, values, thinking, economy, culture, etc.). This includes prophetic statements or dictums (especially references to a “little red book”) but does not include simple assessments or analyses of likely trends.
2. If our nation has had, has, or will have international political weight through diplomacy, economics, military or some other source of power and can thus hope to modify external structures, actions, currents, groups, institutions, etc. (This could be seen as a particular case of #1). Even if it doesn't use it. One example of this is if we put on prior conditions that must be met before we enter into a dialogue with the other party. Another example is "laying down the law" to some other party or issuing threats if it doesn't change its behavior. Another example is picturing a bright national future that depends on no one else but ourselves.
3. If our nation is a model for others or if others are an anti-model for us (without a clear relationship between us and them).
4. If our national future or position we take on issues or national choices matter to others and they have no control over what we do. Including charity we offer to others with no expectation of obligation or repayment on their part. [The distinction between this and deterrence by denial (MC3) is thin. I'd say that you code it WC either if there's no whiff of deterrence in the remark that our future matters to others, or if we are clearly the protagonist in the relationship.]
5. If our nation is ahead in terms of some competitive ranking or measure of national attainments. Plus statements of qualitative superiority and general breast beating as long as it is clearly outward-looking. This includes "it can't happen here" type talk.
6. If our nation intervenes or penetrates any other nation's internal life. N.B. At some point, such penetration if well-established (e.g. holding colonies) could begin to be taken as the status quo, and then any difficulties we ran into in continuing that penetration could be taken as a step down from WC, but not all the way down to "they impact us." Similarly, others holding us as a colony (or part of 'us' in colonial rule) could begin to be taken as the status quo, at which point any difficulties we caused them would be taken as a step down from TC but not all the way to WC (i.e., as MC).
7. Mercantilist model of exchange where we win.

For a coding of “our environment impacts us”:

0. If they are clearly monopolizing the role of protagonist in the relationship.
 1. If the environment influences or penetrates us in any aspect of life, even if we are resisting that influence/penetration. Underscoring this point: calls to resist, counteract, overthrow, undo external presence in our domestic polity should count as “our environment impacts us.”
 2. If our nation lacks international political weight/is facing external structures, actions, currents, groups, institutions, organizations, etc. to which it must conform, to which it turns for redress, or whose bidding it must do.
 3. Submitting to international norms, including as an attempt to or because of anxiety about maintaining or improving our international image or reputation (This could be seen as a particular case of #1). [TC3 vs. WC2: thin line. For me the key is whether we are appealing to norms or to our interests. If we are appealing to norms then we are setting up something external to which we're conforming and we want credit for that. If we are appealing to our own interests then we are setting up our own needs as justification enough, without turning to some external power or authority for justification. If we're suggesting new norms, or bolstering norms such that without our action the norm would go away, however, then that's WC2].
 4. If we are citing an external model/norm/generic case to follow.

5. If our nation is an anti-model for others; if others are a model for us w/o direct relations between the two.
6. If others' national future/behavior/choices will affect us, even if only as a side effect, and we have no control over it.
7. If our nation is behind in terms of some competitive ranking or measure of national attainments. Or general statements of insufficiency.
8. If our nation has membership in a multilateral forum that would exist without its presence. (not just a vague 'club' which actually just means a group of equals, but an actual forum that it has membership in).
9. Mercantilist model of exchange where they win.

Note that the above two are mirror-images of each other for coding.

For a coding of “mutual impact between our environment and ourselves”:

1. If we penetrate/influence the environment and it penetrates/influences us, or if we simply have relations such that neither side monopolizes the role of protagonist.
2. If we faced penetration/influence from the environment in the past and repulsed it.
3. If we are successfully defending against penetration/influence from the outside.
4. If we are facing outside events or inflexible structures that we can't change, but that do not force us to change our preferred course of action significantly.
5. If our (their) actions will affect us (them) yet the other party will still have some influence over the choices that are made.
6. If our nation's rank is exactly the same as some other's in a competitive or numerical ranking or measure.
7. If we claim simple similarity/equality with some other even without an actual relationship. (But here we should not be a mere “case” of an external norm—in that case the coding would be “they impact us”). The difference between this and TC4 is that in TC4 case, there is a relationship—we look at what they do and follow it. In this case, we are both doing the same thing but just because we are in the same position. If the other weren't there and hadn't shown us the way we'd still do the same thing.
8. Statements that would be classed as “we impact the environment,” except that the protagonist is a wider entity of which we consider ourselves an integral part (i.e., that without us, the entity would be fundamentally different if not non-existent).
9. Mercantilist model where neither wins (both knocking the other, or stalemate). Or liberal model of win-win as long as there is true international exchange of goods and services, not just foreign investment by them in us or us in them. In latter case, it's either us penetrating them or them penetrating us.

Note: the coding here represents both the “borderline cases” of the previous coding rules, and also the cases that do not fit those rules.

IV. Coding for Issue-Area (Economic, Domestic, International)

For this coding, the most important rule is the first one: follow the speaker. If the speaker explicitly frames an issue as being "economic," "domestic," or "international," that is how the issue should be coded. For instance, the issue-area of a paragraph that could be interpreted as being about "internal politics" but is found in a section explicitly devoted to "economic policy" should be coded as "economic."

If the issue-area understanding of the speaker is not clear, then we need to impute such an intent. The basic goal of the definitions below is to mirror *traditional* issue-area categories of economic, domestic, and international.

Finally, "domestic" is the broadest, catch-all category. When in doubt, code as "domestic."

*Code as relating to **economic** issues any paragraph⁶⁸ that focuses attention on one or more of the following:*

1. Any statement clearly labeled as being on "economic" issues.
2. Broad statements of economic theories or principles, broad views of the general operation of international and domestic markets.
3. Statements referring to legislation and public administration on issues that are narrowly economic. *"Narrowly economic" is defined as standard topics in macro- or micro-economics, such as the level of income and growth in an economy, unemployment, investment (external and internal), public finance, money supply and banking, international trade, specific economic sectors.*
4. Statements referring to *labor policy* that is limited to fixing wage rates or aim at "flexibility", *science and technology policy* that is merely commercial R&D, *modernization and development policies* whose primary economic effects are direct, *tax and competition policies* justified in narrowly economic terms instead of social justice, *policies toward natural resources* whose value is seen as primarily economic, *public works and infrastructure projects* whose value is seen as primarily economic, *immigration policies* that are primarily framed in a context of the labor supply.
5. Statements referring to the narrowly economic aspects of international treaties or international economic institutions.

*Code as relating to **domestic** issues any paragraph⁶⁹ that focuses attention on one or more of the following:*

1. Any statement clearly labeled as being on "internal or domestic politics." This includes the domestic politics of other states but only if we are not substantially intervening in that domestic situation. If we are then it's international issue.
2. Broad statements of national identity, character, interests, goals.
3. Basic statements of political concepts for ordering society (e.g., rights, justice, order, obligation). This includes global human rights talk that is not clearly tied to intergovernmental treaties or organizations.
4. Statements referring to political institutions (e.g., constitution, regime type, parties) and challenges to those institutions (e.g. political contestation, rebellion).
5. Statements referring to individual national politicians and statesmen.
6. Statements referring to legislation and public administration on issues that are not narrowly economic (e.g., social policy).
7. Statements counted as not narrowly economic would include a statement on *labor policy* that is not merely limited to fixing wage rates, *science and technology policy* that is broader than commercial R&D, *modernization and development policies* whose primary economic effects are indirect, *tax and competition policy* that is justified in terms of social justice rather than merely in terms of narrow economic considerations such as

⁶⁸ Or sub-paragraph, in the rare case that the paragraph has multiple outward-looking references that clearly relate to different issue-areas.

⁶⁹ Or sub-paragraph, in the rare case that the paragraph has multiple outward-looking references that clearly relate to different issue-areas.

efficiency, *natural resources* whose value is seen as broader than monetary value, *immigration policies* that are not primarily framed in terms of the effect on the labor supply.

8. Statements referring to non-political issues in domestic society that are not narrowly economic.
9. Statements referring to non-governmental organizations (such as Amnesty International) or a decentralized "international society" that promote internal behavioral change in non-narrowly economic issues.

Code as relating to **international** issues any paragraph⁷⁰ that focuses attention on one or more of the following:

1. Any statement that is clearly labeled as being on "international" or "foreign policy" issues.
2. Non-narrowly economic references to the structure of and trends in the international system, including norms or rules of international behavior. Also vague references to "world history" or the life of the world.
3. References to a state's non-narrowly economic foreign policies and external behavior, including the actions of the armed forces.
4. Non-narrowly economic references to other states, intergovernmental regimes, alliances, or treaties. This can include references to regional economic integration schemes, if they are not placed in a narrowly economic context.
5. References to international sports competitions or other non-political, non-narrowly economic international interactions.

V. Coding for "Who is Us?"

1. The presumption for each outward-looking paragraph is that "us" is the nation as a whole. The burden of proof is on he who would like to dispute that. Even if a paragraph makes no reference to any "us," the national identity is the default coding.
2. In some cases, nonetheless, the "us" in question might be a wider identity group ("wider"=including the entire nation as a subset). A wider identity can co-exist with a national identity. In coding, if a both a national and a wider identity are expressed with relation to some outward-looking reference, code the wider identity. Unless otherwise noted, the narrower national identity will be assumed to be also held by the speaker. In the case where two wider identities are both expressed with relation to some outward-looking reference, code the widest identity.
3. The key to codings of a wider identity is to understand the definition of a group identity:
 - *An identity is not just a mere category.* In the phrase "the European states system," "Europe" is a geographical expression, not an identity. In the phrase "America is a democratic nation," "democratic nation" is a definition or type, not an identity.
 - *An identity is not just a mere coincidence of values, interests, or strategies.* In the phrase "we have reached an agreement with our GATT partners," we are certainly not expressing hostility toward them, but they are equally certainly "others," not part of our identity.

⁷⁰ Or sub-paragraph, in the rare case that the paragraph has multiple outward-looking references that clearly relate to different issue-areas.

4. A shared identity is a sense of "*we-ness*" (which, among other things, means it's among people, not things). The idea of "*we-ness*" implies *strong (positive) affective ties; a sense of common history and/or a common destiny; an actual or proposed permanent lifting of internal borders beyond the narrow economic sphere; the wider group posited as having interests and values of its own; and the wider group posited as acting or needing to act as a unitary actor beyond its borders*. A clear reference to any one of these aspects would generally lead to a coding of a broader identity.
5. Coding for a wider identity requires subtlety. Some practical clues for shared identity MAY include the following: references to other nations as *members of the "family"* (brothers, sisters, parents, children); references to the wider group's external actions in the singular form ("The EU calls on Israel..."); references to the *individual humans in the wider group without distinguishing their nationality or accepting the primacy of national sovereignty* ("Human rights are for all and are more important than sovereignty"); references to the group as *united or uniting beyond the narrow economic sphere* (and especially political union); *geographical groupings as "isms"* (such as Europeanism, Pan-Americanism...); references to wider group's having *typical symbols of nationhood* ("The EU flag"); sometimes, the use of *capital letters* in denoting the group ("Our Continent must protect itself.").
6. Some clues that a wider group is not an identity group MAY include the following: reference to the wider group *in the plural form* ("the nations of the world are pursuing a solution"); calls for the members of the group to *cooperate or negotiate*; implications that the association is *temporary, limited, or strategic*; discussions of wider group integration explicitly or implicitly *limited to the economic sphere*; discussions of those outside the national group as "*foreign, "external," or "exterior"* or of the national group having normal "*relations*" with those others; if the nation is seen to have *interest or values differences with that wider group*.
7. It is also possible to express an identity that is *narrower than the normal national identity*. For most nation-states a coding of a truly sub-national identity for an outward-looking reference should be rare.⁷¹ Speaking of "our farmers" or "our military" is not expressing a narrower identity, it's just focusing on a subset of the overall identity. However, in some cases, some people who would typically be expected to be included in the national identity may be seen as *traitors* and therefore not members of the nation (e.g., "anti-national elements," "enemies of the people"). This would lead to a coding of "*national identity minus*"-- that is, the identity expressed is the national identity, but not all the people we would generally assume to be covered by that identity are really covered by it.
8. To reiterate: *when in doubt, code the identity as "national."*

⁷¹ Though in divided societies, one can easily imagine that the identity expressed by the "national" leader would in fact be a sub-national ethnic or religious identity.