

What's in a Name? The Effects of World Bank Classifications*

Lindsay Dolan[†]

November 9, 2016

Abstract

Theories of international relations argue that one of the most powerful tools available to international organizations is the ability to classify. But why would a fundamentally discursive tool alter the behavior of powerful elite actors in the global economy? In this paper, I offer a new theory to explain the power of classifications. I argue that classifications, acting as an heuristic device, may directly influence the behavior of some actors in the global economy, but even actors who are less susceptible to cognitive biases may nonetheless use classifications strategically to signal their impartiality. I provide one of the most systematic empirical tests to date of the classification effect by focusing on the dual income classification system produced by the World Bank. Using cross-national data 1987-2015, I exploit the exogenous nature of the thresholds separating categories to show that, consistent with the pattern predicted by my theory, classifications affect the behavior of donors and raters but not investors. I further support my claims with evidence from interviews with a variety of stakeholders in the World Bank. The paper offers new theory and evidence to suggest that international organizations can exercise power through their bureaucracies and are not simply the puppets of powerful state interests.

Keywords: international organizations, bureaucracy, World Bank, aid

*I thank Chris Blattman, Justin Canfil, Allison Carnegie, Michael Clemens, Alicia Cooperman, Alan Gelb, Kolby Hanson, Scott Morris, Todd Moss, and Tara Slough, as well as participants in workshops at Columbia University and the Center for Global Development, for helpful comments and conversations. I am grateful to the Center for Global Development for the fellowship that enabled this research. Jamie Park provided helpful research assistance. All remaining errors are my own. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship, grant number DGE-16-44869.

[†]Ph.D. Candidate, Political Science, Columbia University. E-mail: lindsay.dolan@columbia.edu

By 1989, World Bank economists had collected the first decade of the World Development Indicators and had at their fingertips the most complete data to date on global poverty and development. Like many researchers, their first goal was to present illustrative patterns and descriptive statistics to guide policymakers at the World Bank in their efforts to address poverty. Up until this point, these analysts had grouped countries into “developing,” “industrialized,” and “oil-exporting” economies, but the categories were poorly defined and largely ad hoc. Hoping to arrive at something more systematic to help them characterize these trends, the economists introduced an income classification system separating countries into low, lower-middle, upper-middle, and high income country categories. Although they never intended the categories to be used for operational purposes within the World Bank or to be used by anyone outside the World Bank, today this same classification system is widely used in the global development landscape.¹

These World Bank researchers unwittingly participated in what Michael Barnett and Martha Finnemore, in their seminal article, have characterized as “one of bureaucracy’s greatest sources of power:... the ability to classify objects.”² For Barnett and Finnemore, classification offers a means by which the bureaucracies of international organizations can extend the power of those organizations. For example, they argue that the United Nations High Commissioner for Refugees expanded its power by using its authority as an expert in humanitarian affairs to expand the initially stringent definition of a “refugee” to justify its action in more cases.³ By mapping the social world through labels, these bureaucracies not only justify (and expand) their existence; they also provide a vocabulary and partitioning that other international actors may choose to use.⁴

The landscape of development is littered with these classifications. States may belong to the Least Developed Country (LDC) category, the Fragile and Conflict Affected State (FCAS) category, the Highly Indebted Poor Country (HIPC) category, and others, in addition to the World Bank income classifications described above. Each of these categories is created and maintained by a prominent international institution that devotes valuable staff and resources toward ensuring that the category remains fit for purpose. Indeed, many other international actors often contest the grounds for categorization. For example, the Raise the MIC campaign lobbies the World Bank to raise the ceiling separating low income and middle-income countries because the current system, in their view, “understates the true extreme of poverty in the MICs”⁵ and

¹See Nielsen (2011); Fantom and Serajuddin (2016).

²Barnett and Finnemore (1999, 711).

³Barnett and Finnemore (2004, ch. 5).

⁴Scott (1998) makes a similar argument to explain the power of state bureaucracies at the domestic level. In his view, states become powerful by making their populations “legible” to the state through censuses, mapping, and records.

⁵<http://raisethemic.org/>, accessed September 12, 2016.

disadvantages them with respect to obtaining international assistance.

Yet the production of these categories has received scant attention in studies of global development institutions. Far from aiming to understand the power that lies within the bureaucracy of these global development institutions, the literature on the World Bank and International Monetary Fund tends to instead characterize these institutions as mere puppets of the powerful state actors that founded and support them.⁶ While the literature presents compelling evidence of “politics at work” in the disbursement of aid and lending to recipients with greater representation on the Boards of these institutions, the fixation on these distributive politics neglects the political power these institutions derive *independent* of their funders and backers. Indeed, influential observers of the World Bank argue that today, particularly in a low interest rate environment with new alternative sources of finance, the Bank’s primary role is not in lending but rather in facilitating international agreements and providing expertise and technical assistance.⁷ Yet we know little about the tools that the World Bank uses to achieve these objectives.

In this study, I focus on classification as an instrument by which the bureaucracy of the World Bank shapes the behavior of other actors in the international economy. In brief, I show that classifications matter; more fully, I offer a theory that explains why they matter, predictions that suggest under what conditions they matter, and evidence that supports the pattern predicted by the theory. My cognitive-strategic theory suggests that these World Bank classifications are likely to influence the behavior of donors and rating institutions but not of investors, and I find strong support for my theory in cross-national data and in qualitative interviews. Becoming a “lower-middle income country” in particular causes countries to steeply lose access to aid but improve their reputations for being creditworthy, and becoming a “low income country” causes countries to be perceived as more democratic. In demonstrating the effects of classification, I provide the most systematic empirical test to date of Barnett and Finnemore’s claim that classifying objects is a critical source of power for international organizations. By reintroducing bureaucratic tools into the study of international organizations, I offer an alternative account of their power.

This study enters a growing literature on the importance of labels in political science. For example, previous studies have found that investors’ risk perceptions are affected by a country’s club membership

⁶For examples of empirical works on distributive politics in development institutions, see Kuziemko and Werker (2006); Dreher et al. (2009); Kaja and Werker (2010); Kersting and Kilby (2016). More broadly, the rational design approach to international organizations argues that “states use international institutions to further their own goals, and they design institutions accordingly” (Koremenos et al. 2001). For this reason, prominent theorists have often dismissed international organizations as mere reflections of an underlying distribution of power (Waltz 1979; Mearsheimer 1994).

⁷See Ravallion (2016); Clemens and Kremer (2016).

and by the other peer countries with which it is grouped.⁸ In the domain of global governance, ratings and rankings organizations have proliferated in recent decades and routinely evaluate countries on their corruption, human rights records, level of democracy, and other dimensions. Others have shown that the graded countries, perceiving the benefits of desirable labels, often adopt policies targeted at changing their scores or petition the ratings organizations to amend them.⁹ Despite the growing awareness that labels matter and invite political contestation, we have little theoretical understanding of why and when they matter. In addition to extending the study of labels to the development context, my project builds a theoretical framework in which we can make sense of these findings.

Many of the existing case studies of labels, ratings, and indices focus on those instruments that have been created by advocacy groups hoping to effect policy change, but by focusing on the World Bank classifications, I show that these effects hold even in a system that was not devised for political purposes. My study therefore also calls attention to the unintended consequences that can result from creating these metrics. Whether for activist or benign purposes, many organizations produce classification systems and frequently use arbitrary thresholds to separate categories. These thresholds may lack meaning initially, but I show that they can acquire significance over time. This illustration of the power of path dependence has important consequences for millions living in poverty on each side of these thresholds.¹⁰

In what follows, I offer a general cognitive-strategic theory of classification and derive implications for which classifications are most likely to influence the behavior of which actors in the international community. Next, I narrow my focus in this paper to the World Bank classification system. I describe the context and offer specific, testable hypotheses that relate to my theory. Making use of the arbitrary nature of the cutoffs that define these classifications, I test the hypotheses quantitatively on cross-national data spanning the period 1987-2015. My results demonstrate that, consistent with my hypotheses, the classifications affect the behavior of donors and raters but not investors. To further support this argument, I also present evidence from qualitative interviews with various stakeholders in the World Bank, whose experiences align with the patterns I find in the data.

⁸See Gray (2013); Brooks et al. (2015).

⁹See, for example, Cooley and Snyder (2015); Kelley and Simmons (2015).

¹⁰See David (1985); Pierson (2000).

Theory

Why would we expect classifications produced by an international organization to influence the behavior of other actors in the international economy? In this section, I offer a cognitive-strategic theory to account for the disproportionate power these classifications acquire. In brief, I lay out a cognitive mechanism by which classifications act as heuristic devices to simplify decision-making, and a strategic mechanism by which actors use classifications produced by a third-party actor to demonstrate the impartiality of their decisions. I show that the two mechanisms interact. Finally, I contrast my theory with a conventional explanation for the power of international organizations, which focuses on the role of information.

Cognitive-Strategic Explanation

Extensive research in behavioral economics has documented that many decision-makers are susceptible to cognitive biases that sway their decisions in irrational ways. One important finding from this literature is that human brains are hard-wired to find shortcuts, known as heuristics, that help them to process overwhelming information.¹¹ Like a computer, human brains are constantly applying algorithms to distinguish relevant from irrelevant information, but this is a very cognitively demanding task. Moreover, we experience deep discomfort, termed “cognitive dissonance,” when we encounter conflicting information, and look for shortcuts to avoid grappling with every data point we encounter. Categorizations offer one such shortcut: a decision-maker can make inferences about an unfamiliar case based on her familiarity with other cases in the same category.¹² In the domain of political science, others have shown that peer country groupings influence sovereign lending practices.¹³ Given the amount of contradictory and poor-quality data in the development landscape, classifications could be very attractive means of simplifying a messy picture of countries on the cusp of development.¹⁴ In other words, this explanation argues that classifications are so powerful because they distill information and, in so doing, reduce the cognitive dissonance that is so anathema to humans.

A related phenomenon in cognitive psychology called the “halo effect” could amplify the effects of classifications. Social psychologists have amassed evidence that an individual’s opinions about an actor’s attribute A are influenced by information about that actor on unrelated attribute B. For example, evidence shows that people who are thought to be more attractive are often perceived as being more competent or

¹¹See Kahneman (2011).

¹²See Taylor (1981).

¹³See, for example, Brooks et al. (2015); Gray (2013).

¹⁴Jerven (2013) documents the extent of the unreliability of data in the developing world.

intelligent.¹⁵ If actors in the international community are also susceptible to this bias, classifications could affect international perceptions on more dimensions than just that which they seek to measure. For instance, countries that graduate in an income classification might also more easily persuade observers that they are democratic or respect human rights.

An important requirement to a good heuristic device is that the observer is familiar with the classification system. Taylor (1981) notes that the propensity to stereotype based on group affiliation depends on “the extent to which a perceiver has a well-developed concept for that attribute. ... A Korean is simply [Asian], if one is unable to distinguish Koreans from [Asians]” (86).¹⁶ As a result, we would expect only classification systems that are intuitive to the observer to exert effects.

However, an important constraint to the explanatory power of the cognitive explanation on its own concerns the decision-making processes used by internationally important actors. Cognitive psychology may be able to account for stereotypes or implicit biases among the mass public, but there are two reasons to expect them to be less powerful among geopolitically important actors. First, these actors are often elites, who may have more informed priors about the involved cases that reduce their sensitivity to cognitive biases. Second, decisions made by organizations, firms, and countries are filtered through a more institutionalized and regulated process that may temper the effects of cognitive biases.¹⁷ Psychological research shows that asking individuals to justify their treatment of cases reduces the effects of implicit bias.¹⁸ Actors like donors and investors must constantly propose and defend budgets to a variety of individuals with the opportunity to veto. Even if elite individuals are vulnerable to these biases, limiting their individual discretion over a final outcome will mitigate the opportunity for bias to sway the decision.

What then explains the sensitivity of these elite and institutionalized actors to classifications? I argue that these actors are behaving strategically. Many of these actors are responsible for making contentious distributive decisions that will produce backlash from those who are adversely affected. When they want to avoid being perceived as partisan or political, it is strategic for them to rely on objective criteria produced by third parties to justify their decisions. Doing so allows them to claim that their hands are tied and avoid

¹⁵For a review, see Greenwald and Banaji (1995).

¹⁶Taylor (1981)

¹⁷Indeed, one of the most prevalent critiques of psychological theories of international relations is that these theories rarely offer an argument for how individual-level psychological traits explain behavior at the state level-of-analysis. For example, in their critique of the research on status inconsistency Dafoe et al. (2014) argue, “IR studies that used this as their theoretical foundation did not typically contain any theory that explained why we should expect individual-level results linking inconsistency to “violent” or “dysfunctional” behavior to translate directly to world politics.” (388) Dafoe et al. (2014)

¹⁸See Taylor (1981).

taking responsibility for allocational decisions. This logic should sound familiar to students of delegation and agencies. Weaver (1986) notes that one way legislators often avoid blame is to “cede discretion to the president or an independent agency for making politically costly decisions” (375).¹⁹

Classification systems can act as independent agencies in the sense that they can be used as an apparently third-party, impartial rubric for actors who are allocating scarce goods. If an organization develops policies that treat countries differently on the basis of their classification by a third party, that organization no longer needs to be responsible for arriving at its own eligibility scheme and risk political backlash, either from those who lose in the system, or from principals who question the motives of their agents. Creating classification systems is politically costly, so where possible, organizations have an incentive to use those created by others.²⁰

Which actors are the most likely to defer to external classifications in their own operations? I argue that the actors most likely to behave in this way are those with a large demand for the appearance of impartiality. Not every actor needs to appear impartial. Impartiality is likely not important for private investors, who only need to show their backers a return on investment. Among donors, there is likely some variability in the demand for impartiality. While some explicitly use aid as a political instrument, others tout their impartiality. The Millennium Challenge Corporation (MCC), for example, was commissioned to be the apolitical aid instrument for the U.S. government, so it has a very high demand for impartiality. It is therefore no surprise that MCC has developed an eligibility and allocation rubric that relies heavily on democracy measures produced by Freedom House, an external agency.²¹ Whenever MCC’s allocations are contested, it can simply reply that its clients’ issues are with Freedom House, not with MCC.

If organizations make use of classification to appear impartial, whose opinions are they concerned with changing? The specific answer to this question will depend on the organization in question and may include a number of different audiences, but for the purposes of this theory, it is sufficient to focus on the common characteristics of those audiences that organizations are looking to impress. The most important common characteristic is that these audiences will tend to be less elite and have less institutionalized decision-making procedures than the decision-makers in the organization my theory is concerned with. For example, one of the important audiences who may demand impartiality from a donor is the domestic tax base who funds

¹⁹For arguments about delegation to independent agencies as a means of depoliticization, see also Flinders and Buller (2006); Landwehr and Böhm (2011).

²⁰The World Bank data group has received much criticism regarding its classification system. See Fantom and Serajuddin (2016).

²¹Working paper obtained through private correspondence with author. Sarah Bush, “The Politics of Rating Freedom: Ideological Affinity, Private Authority, and the Freedom in the World Rankings.”

foreign aid programs.²² A domestic taxpayer will be both less informed about development and is unlikely to ever engage in a rigorous evaluation process. The general pattern is likely to hold across any organization that is accountable for appearing impartial to some principal—whether it is a median taxpayer, representative, or delegate—because of the nature of principal-agent relationships. When we think about the organization as the agent of some principal who expresses a preference for impartial treatment of countries, the agent is tasked with this goal precisely because it is viewed as having expert judgment and information, or because it has the time and resources to institutionalize a formal decision-making procedure. The audiences or principals, then, that may be imposing a criterion of impartiality, will nearly always be less well informed and less rigorous in their evaluation process.

Because of this, cognitive biases once again become relevant for my theory. Recall that the actors that are most susceptible to cognitive biases are less elite with less institutionalized decision-making processes. This describes most of the audiences that international political and economic need to project an appearance of impartiality to. As a result, even if the major economic actors are not as prone to these biases, they may strategically adjust their behavior to appear impartial to those who are susceptible to these biases.

I summarize by discussing the observable implications that result from this cognitive-strategic theory. (In the next section, I will connect these observable implications to testable hypotheses that are specific to my case.)

Implication 1. Classifications are likely to have larger effects on actors with less institutionalized decision-making processes. The more directly that perceptions can map on to outcomes, the greater the role for cognitive biases to operate directly.

Implication 2. Classifications are likely to have larger effects on actors with greater demand for impartiality. When actors want to avoid responsibility for politicizing their decisions, they will be likely to defer to classifications produced by an independent actor. Since impartiality is judged by actors more susceptible to cognitive biases, cognitive biases operate indirectly.

Implication 3. Simpler, familiar classifications are likely to exert stronger effects than more complex, expert classifications. This is true in any relationship in which cognitive biases are at work.

²²Milner (2006) uses this logic to explain that bilateral donors will finance multilateral agencies, who are more trusted by the mass public to carry out development, in an effort to reassure their funders that they are providing needs-based aid.

Informational Explanation

I have claimed that classifications can acquire undue significance for some audiences because of a cognitive desire for heuristics, and that in turn this can shape the strategic calculations of actors who want to avoid the appearance of politicization. But a simpler explanation could be that classifications are actually communicating valuable information that observers otherwise lack. This accords with a long tradition in political science that focuses on the informational role of international institutions. Keohane (1984), for example, argues that reducing information asymmetries between international actors is one of the main reasons that states agree to cooperate in regimes. It is possible, then, that classifications derive their power from reducing information asymmetries in the international community by communicating information about classified countries. More disadvantaged international actors may interpret classifications as communicating valuable information and utilize these classifications in their decision-making processes.

One observable implication of this explanation would be that more complex classifications are likely to have more impact. A classification system could be made complex by incorporating data that is not available to the public or by using subjective assessments made by in-house experts. However, even a classification system using public data could be complex if it synthesizes many kinds of data by creating an index, which involves the imposition of cutoffs or weights. As social scientists, we understand that it is an expert skill to construct accurate measures; international actors know this too, and may trust international organizations to hire experts who will craft a classification system that maps on to its object of measurement. A simple classification system with little logic embedded in its formulae, however, is unlikely to be very influential. This implication differs from my own expectation, which is that the simpler classifications will be most powerful.

A second observable implication of this explanation would be that classifications should have larger effects on observers who suffer most from an informational asymmetry. Actors who are themselves expert in similar topics to the classification system in question should find less use in relying on the classification, whereas actors who lack this expertise will find it more valuable. I expect this implication runs in the opposite direction of my own: the actors most likely to require the appearance of impartiality are those with access to sufficient information to make self-serving decisions. Given these contrasting patterns, I can proceed to empirically test my own cognitive-strategic explanation against this competing account grounded in informational asymmetries.

The World Bank

In the domain of development, many actors try to classify countries according to their level of development, and arguably the most influential actor is the World Bank. The World Bank was founded in 1944 in Bretton Woods, New Hampshire, alongside the International Monetary Fund, and it was tasked with long-term development. In particular, the World Bank was founded to correct market failures and provide developing countries with access to finance in the form of subsidized loans. The two main lending arms of the bank are the International Bank for Reconstruction and Development (IBRD), which was created at the World Bank's inception to help re-build Europe after World War II. It continues to provide developing countries with flexible loans at favorable rates, but they are non-concessional, meaning that the IBRD is financially self-sustaining and does not rely on donors to subsidize borrowing by its members. However, many countries were unable to borrow on these terms. The International Development Association (IDA) was founded in 1960 to provide concessional finance and grants to countries who were not creditworthy enough to borrow from IBRD. IDA is primarily financed through replenishments from donor countries every three years and serves the poorest of the World Bank's clients. As new sources of finance for developing countries have emerged, many have argued that the contemporary World Bank should primarily be seen as a "knowledge bank" that can primarily offer technical assistance and support contributions to global public goods.²³

In its lending policies and in its analysis, the World Bank produces two different classification systems. Several features make these classifications an attractive object of study. First, they are substantively important and a salient vocabulary used by the broader development community. Second, that there are multiple classification systems allows me to study variation in the characteristics of each classification system. Third, both classification systems make use of a series of arbitrary thresholds to separate groups within each system. As will be discussed later, this introduces helpful exogeneity and expands the opportunities for causal identification. Fourth, because the classifications are predicated on a relatively volatile variable – national income – there is a great deal of movement across threshold boundaries. In other classification systems, such as the LDC system produced by the UN, there are very few "graduations," limiting the ability to document an empirical relationship between a country's classification and the outcomes it experiences.

Fifth, global changes in the demography of poverty have attracted much attention to both the World Bank's analytical framework for classifying countries as well as its policies for "graduating" countries from

²³See Ravallion (2016); Clemens and Kremer (2016)

IDA to IBRD loans. The world's poor increasingly live in countries that have experienced rapid growth rates in recent years and are graduating from LIC status and subsequently crossing the operational cutoff.²⁴ This increases the policy relevance of my research (it is important to understand what political consequences these transitions will entail) as well as the evidence available to test my hypotheses (many actors are engaged in a vocal debate about if and how these systems should be reformed).

Classification Systems

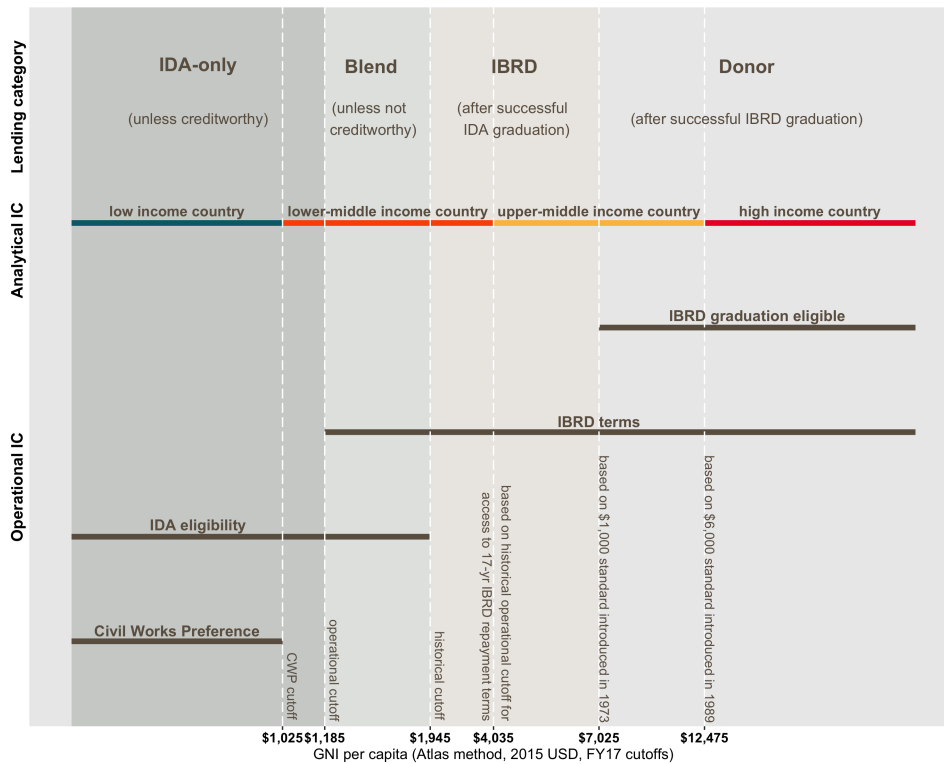
Although they are commonly conflated, the World Bank actually produces two income classifications: analytical and operational classifications. Both classifications sort countries into four groups by calculating whether the country's gross national income (GNI) per capita, a measure of a country's overall income divided by its population, falls above or below certain pre-defined thresholds separating categories.²⁵ Although the two systems share some thresholds, some other thresholds differ, and the two systems are used for different purposes. Only operational classifications are used for determining whether a country will be able to access finance from IDA, IBRD, or both. In contrast, the analytical classifications are produced by the World Bank for analytical purposes, but have no bearing on World Bank policy. It is therefore a widespread misconception that "low-income countries" (an analytical grouping) are synonymous with IDA recipients, when in fact there are more IDA-eligible countries that are "lower-middle income" than "low-income."

The analytical classification system labels countries as "low income," "lower-middle income," "upper-middle income," or "high income" (LIC/LMIC/UMIC/HIC). These classifications are presented in color in the middle of Figure 1, and can be easily calculated by determining where a country's GNI per capita is relative to a series of thresholds. Every year on July 1, the World Bank Development Economics Group publishes its new classifications, which are made on the basis of GNI data from the previous calendar year

²⁴See Moss and Leo (2011); Morris and Gleave (2015).

²⁵The measure that determines a country's classification in both systems is GNI per capita, valued annually in US dollars using the Atlas method. The Atlas method uses an average of the current year's exchange rate and the previous two years' exchange rates inflated to current year prices, and is employed to reduce volatility of the estimation of GNI per capita. To calculate exchange rates, the Atlas method compares a country's prices to a weighted average of prices from the economies who have Special Drawing Rights in the IMF. This is known as an "SDR deflator." Several objections have been raised regarding the calculation of these income data and the selection of this variable as the basis for income classification. In general, skeptics contend that the current measure is too blunt to meaningfully capture development, while its defenders argue that it is the only indicator consistently and accurately reported by nearly all countries. The more indicators and adjustments incorporated into the calculation or classification of income, the less universal and comparable the results will be. For example, one levied critique is that using GNI data without adjusting for purchasing power parity (PPP, or how much a dollar buys in each economy) fails to take into account important variation in price levels between countries. This concern has been raised repeatedly both to both World Bank data technocrats and also to the World Bank's Board of Executive Directors, but both groups have each time concluded that the coverage and quality of price data is insufficient to produce a robust cross-national data set of income. See Fantom and Serajuddin (2016).

Figure 1: Income classifications and lending categorizations in the World Bank



and on the historical thresholds, which are updated only to account for inflation. These classifications were never intended to be and have never been used for operational purposes.

For operational decisions, the World Bank uses a different set of classifications, shown in black lines in the bottom half of Figure 1. These classifications are an important but not the sole determinant of whether countries may borrow from IDA, IBRD, or both. The most important operational cutoff for my study is the one that triggers the multi-year process of graduating from IDA, and it is simply referred to as the “operational cutoff.” When a country crosses this cutoff, the country desk will request an assessment of the country’s creditworthiness by the IBRD to determine whether the country will be able to access IBRD loans. If it is found creditworthy, it becomes reclassified as a “Blend” country for the following fiscal year and can access both IDA and IBRD loans.²⁶ After a country has crossed the operational cutoff for three consecutive years and has been found creditworthy by the IBRD, it becomes reviewed for graduation from IDA during the next replenishment meeting, a meeting that occurs every three years in which donors pledge to replenish the stock of IDA that will be lent to countries in the next cycle.

²⁶However, the terms on its old IDA credits will harden. Also, if it is not found creditworthy, it becomes a “gap” country: the country may still access IDA on hardened terms, but cannot yet access IBRD.

As is apparent from this discussion, income is an important criterion in the type of assistance a country receives from the World Bank, but it is not the only criterion. Creditworthiness, for instance, also plays an important role in a country's ability to borrow from IBRD. Other criteria appear when a country is evaluated for graduating from IDA. Because graduation from IDA is very costly for states²⁷, and because the institution wants to avoid reverse graduations, IDA staff or deputies frequently recommend against graduation if they anticipate political or economic instability for any number of reasons.²⁸ The result of these negotiations is that, in practice, countries spend an average of six years in Blend status (i.e. borrowing from both IDA and IBRD) before they are graduated entirely from IDA.

Thresholds

If these thresholds are so significant for separating countries into categories, then where do they come from? Many have investigated the historical origins of these thresholds and have found that they were arbitrarily selected for reasons that are irrelevant to today's world.²⁹ Recall from the introduction that the analytical classifications were born of the need for a simple analytical instrument to track development progress. When formalizing this system in 1989, researchers selected several figures from outdated World Bank policies to separate categories. The LIC ceiling, distinguishing LICs from LMICs, is actually based on a cutoff introduced in the 1970s called the "civil works preference" cutoff, below which countries received preferences in civil works procurement bids in Bank-financed projects because they were not thought to be competitive enough. The threshold separating LMICs from UMICs, was based on a different operational threshold no longer in use: the cutoff used by the IBRD to assess 15-year versus 17-year repayment terms, categories that have since been collapsed.³⁰ As for operational policy, the most important cutoff has always been the one determining when countries would begin the process of graduating from IDA. The first operational cutoff was introduced in 1964 at \$250 per capita, but by 1989, IDA's resources were too limited to accommodate all countries falling below this cutoff. IDA lowered the operational cutoff by 38%, which was \$580 per

²⁷They lose access to future IDA credits, often face accelerated repayments on old IDA credits, and borrowing limits in IBRD sometimes mean that the total volume of borrowing will be constrained.

²⁸In the most recent graduation policy review document, IDA staff supported its recommendation for graduating three of the twelve current blend countries by citing the following indicators: poverty headcount ratio, Human Development Index, real output growth, nominal public debt, commodity exports, political stability and absence of violence/terrorism, risk of debt distress, Worldwide Governance Indicators, and Economic Vulnerability Index. In case descriptions, the authors mention low prices on commodity exports and lack of market access. World Bank IDA Resource Mobilization Department (2016)

²⁹See Nielsen (2011); Knack et al. (2014); Fantom and Serajuddin (2016).

³⁰While I don't focus on it in this study, the UMIC/HIC threshold was simply set at \$6,000, and no rationale has ever been found for this number.

capita in US dollars. The old operational cutoff became known as the “historical cutoff.” Although it has no operational significance anymore, IDA graduation reviews often note the number of years a country has exceeded the historical cutoff when recommending graduation. As such, the historical cutoff is informally relevant for IDA-eligibility. In other words, when states cross the “operational cutoff,” this triggers the process of graduation, and when states cross the “historical cutoff,” this makes the need for graduating them more urgent. Both of these operational cutoffs, however, are not related to any current operational policy, and if any strategy contributed to their selection, it does not systematically separate countries any longer.

While the origin of each threshold is different, their shared characteristic is that they are all historical artefacts that make them exogenous to today’s development environment. As such, while a country’s GNI per capita is not random, whether its GNI per capita falls on one side of a threshold or another is. This makes it possible to devise an empirical strategy that will identify the effect of the classification itself when we control for a country’s income. In this study, I will use the LIC ceiling and the LMIC ceiling to study the effects of the analytical classification system and the operational and historical cutoffs to study the effects of the operational classification system.³¹

Testable Hypotheses

Recall that the most important implications of my theory concern the variety in effects we would expect to observe across classification systems and across international actors. Having discussed the context of the World Bank in development politics, I am now able to map these implications onto testable hypotheses specific to this context. In this paper I will look at the variety in effects *of* the two different classification systems produced by the World Bank as well as the variety in effects *on* the behavior of important actors in the international political economy: donors, private investors, and raters. To introduce additional variation, I will investigate the effects on multiple types of donors – traditional bilateral donors, emerging bilateral donors, and multilateral donors – as well as different types of raters – creditworthiness raters and democracy raters.

In order to test the first observable implication, which suggests that classifications will affect actors without institutionalized decision-making processes that mitigate the power of cognitive biases, I first characterize these actors in terms of the extent of institutionalization in their decision-making process. The most

³¹I omit some higher cutoffs—namely, the UMIC ceiling and the IBRD graduation threshold—to preserve a focus on how classification affects developing countries. For a study of the operational threshold determining a country’s eligibility to graduate from the IBRD to donor status, see Carnegie and Samii (2016).

institutionalized actors here are donors and investors, who must propose and defend budgets to a variety of constituencies who have an opportunity to veto or weigh in on policy decisions. This process reduces the opportunities for cognitive biases to influence behavior, as actors are forced to defend their proposals on rational grounds. On the other end of the spectrum, raters are subject to very few institutionalized decision-making processes. Coding judgments involve a baseline level of subjectivity, and while most credible ratings institutions adopt steps designed to minimize arbitrary evaluations, some subjectivity will always remain. That said, there is variation between raters based on the number of steps taken to improve the accuracy and objectivity of the grading process. For example, ratings that are based on private and anonymous surveys will be the least institutionalized sources. These actors will be the most susceptible to cognitive biases. Therefore, if only the cognitive mechanism is operating, I expect thresholds to exert a greater effect on ratings than on aid or investment.

The second observable implication focuses on an actor's demand for impartiality, which I deduce by looking at the principal-agent relationships each of my actors is involved in. Who is each accountable to, and does that principal demand impartiality? By this logic, I expect donors to have the greatest demand for impartiality because they are tasked with distributing valuable goods, in principle at least, for the benefit of the recipient. Others argue that principals – be they domestic taxpayers or legislatures – hold aid agencies accountable for engaging in good development practices because they are acting as surrogates for those who need development most, but have no formal means of holding donors accountable.³² Among donors, too, I expect to see that multilateral donors have a greater demand for impartiality than bilateral donors because multilateral donors are accountable to boards representing a variety of national interests, whereas bilateral donors may have greater license to incorporate their own national interest into distributional decisions. The appearance of impartiality is not necessary for private investors, whose incentives are to identify investment opportunities with the greatest returns, not to satisfy some distributional or fairness preferences of a principal that they are accountable to. Although ratings agencies strongly desire the appearance of impartiality, they have a greater ability to provide that impartiality for themselves, and in fact, it would undermine their credibility as a ratings organization to so obviously defer to the ratings of a different organization. Consequently, both private investors and ratings organizations are coded as having a low demand for impartiality, and are therefore the least likely actors to exhibit a classification effect according to the purely strategic

³²See Martens (2005); Hawkins et al. (2006).

mechanism in my theory.³³

When we combine these mechanisms, my theory expects that I will observe effects on raters and aid, but not investment decisions. While raters lack a strategic reason to use classifications, we will be more likely to observe cognitive biases at work directly. In contrast, while donors are unlikely to reflect cognitive biases in their behavior, they may involve classifications strategically in their behaviors to gain the support of those who are susceptible to the cognitive classification effect. But investors have no such strategic reason, nor are they likely to exhibit a cognitive bias. This results in the following hypothesis.

Hypothesis 1. Thresholds are more likely to affect raters [cognitive mechanism] and aid [strategic mechanism] than they are to affect investment decisions.

The third observable implication concerns the difference in the type of classification. Since cognitive psychology is at work both directly and also indirectly through the strategic explanation, I expect that simpler, more familiar classifications will exert the strongest effects. In the context of the two classifications under study, this means that the more widely-known analytical classification system is more likely than the operational classification system to affect the outcomes described in Hypothesis 1. Not only is the analytical system more widely used, but it is also simpler and more intuitive to understand, whereas a user of the operational classification system must be familiar with the different lending arms of the World Bank and the criteria for graduation from one to the other.

Hypothesis 2. Analytical thresholds are more likely to have an effect than operational thresholds.

Expert Priors

While my theory predicts which classifications influence what actors, it does not actually predict the direction of the effects. To arrive at directional predictions would demand developing additional theories for the costs and benefits each actor faces for allocating aid or investment to one country over another, or rating a country a certain way.³⁴ As such, I don't develop predictions for the direction of effects but just look for where the effects occur.

Nonetheless, prior to conducting my analysis, I conducted a brief survey of experts at the Center for Global Development to solicit their priors not only on where the effects would be but also on the direction.

³³Others have already documented evidence of donor herding, although they do not offer theories to explain this behavior. See Knack et al. (2014). Also see Frot and Santiso (2011).

³⁴In fact, Knack et al. (2014) begins with competing hypotheses regarding donor behavior: donors could add to the countries they observe other donors funding, or they could do the opposite in an attempt to compensate for the behavior of others.

CGD is one of the most actively involved think tanks in IDA graduation policy reform and many of the fellows have decades of experience in multilateral organizations working to promote development through aid, investment, finance, and governance initiatives.³⁵ To sufficiently define my research question for them, I presented respondents with a list of the outcomes I would be examining, potential explanations for any effects, and potential explanations for null effects.

The results appear in Table 1. First, the surveyed experts expressed mostly uniform beliefs about the direction of any potential effect of crossing a threshold. The group tended to believe that threshold crossings would decrease aid, increase investment, and improve credit ratings, democracy ratings, and the incumbent's chances of being re-elected. This provides me with a helpful basis for judging whether any of my effects are surprising or counter-intuitive.

Second, there was substantial variation in beliefs about which classifications would produce effects, and which outcomes would be affected. Taken together, my respondents thought classification effects were more likely to exist for aid and FDI than for ratings and re-election probabilities. They also thought classification effects were more likely to exist on the operational category change (graduating from IDA) rather than the analytical one (crossing the LIC ceiling).³⁶ But even within each outcome and each classification system, there was significant variation among experts in their beliefs about which classifications would matter and whose behaviors they would affect. At the very least, this underlines that the expected effects of classifications are not obvious and that better theories and evidence are needed to understand this phenomenon.

Table 1: Survey of Priors from CGD

Expected direction of effect:	Crossing LIC ceiling			Graduating from IDA		
	-	+	null	-	+	null
Aid	10 (.59)	1 (.06)	6 (.35)	14 (.82)	0	3 (.18)
FDI	0	12 (.71)	5 (.29)	1 (.06)	13 (.76)	3 (.18)
Credit rating	0	8 (.47)	9 (.53)	0	12 (.71)	5 (.29)
Other ratings (e.g. democracy)	0	4 (.24)	13 (.76)	0	5 (.29)	12 (.71)
Prob. of incumbent re-election	0	4 (.24)	13 (.76)	1 (.06)	6 (.35)	10 (.59)

Notes: N=17. Table reports count, with frequency in parentheses.

³⁵While all staff were invited to participate in the survey, I subset my sample to only my 17 respondents who reported 3 or above on a 5-point scale when asked about their expertise/familiarity with IDA graduation policy.

³⁶This survey question is not exactly comparable to my results, since I asked experts about the category change from IDA-only or Blend to IBRD-only status, not about crossing the operational cutoff. Nonetheless, it is a good proxy for the perceived importance of operational categories.

Data

To study the effects of classifications, I constructed a data set containing a variety of economic data and classifications during the period 1987-2015. I defined my sample as all country-years that ever received an analytical income classification from the World Bank, but since I am interested in countries who cross the lower threshold, I exclude all countries that have been continually classified as HICs since before 1989. Because the present categories were defined in FY89, the first observation for most countries therefore contains the classification awarded in FY89 (July 1, 1988-June 30, 1989), which was determined in spring 1988 on the basis of data from calendar year 1987.³⁷

The main explanatory variables are historical data on gross national income (GNI) per capita and the historical thresholds that were used to classify countries in each year. I obtained the original GNI data that were used to classify countries at their time of classification from the World Bank Development Economics Data Group.³⁸ It is important to note that this figure can differ dramatically from the estimate of GNI that can be obtained by downloading the most current World Development Indicators online. This is because income estimates can be revised over time on the basis of updated economic assumptions, new population data, and other causes. Because I use the historical income data and the historical thresholds, this is the best way of approximating how a country was perceived at the time of its classification, regardless of how it would be classified ex post on the basis of contemporary data. The historical thresholds are available online from the World Bank.³⁹

I collect a variety of outcomes to represent the responses of various actors in the international community. To capture the classification effect on donors, I retrieved data on aid disbursements from the OECD. In keeping with the literature, I looked at net disbursements of official development assistance (ODA) and imputed zeroes for missing values.⁴⁰ I aggregated bilateral donors by various types, namely, whether they are members of the OECD's Development Assistance Committee (DAC). DAC donors tend to be traditional donors, including many from Europe, while non-DAC donors are widely seen as new or emerging donors, such as China and Saudi Arabia. Separately, I also aggregated multilateral donors, such as the African

³⁷Unless otherwise indicated, years are reported in terms of data years.

³⁸The World Bank provided the GNI data used for the analytical classifications from 1999-2015 (i.e., used to classify economies for FY01 onwards) and the GNI data used for the operational classifications from 1973-2007. In this paper, I use the analytical figure for all years after 1999 and the operational figure for years before. Where they overlap, the numbers are very close to one another.

³⁹<http://databank.worldbank.org/data/download/site-content/OGHIST.xls>

⁴⁰See Knack et al. (2014) for precedent.

Development Bank and GAVI (the Vaccine Alliance).

To capture the effect of the classification on private investors, I used data on foreign direct investors. Following convention, I incorporated data on net annual inward FDI flows as a percentage of GDP from UNCTAD's Handbook of Statistics.⁴¹

To capture the classification effect on raters of creditworthiness, I gathered the biannual country risk ratings published in the investment journal *Institutional Investor*. According to the magazine, the IIR ratings "are based on information provided by senior economists and sovereign-risk analysts at leading global banks and money management and securities firms. They have graded each country on a scale of zero to 100, with 100 representing those countries that have the least chance of default. ... The individual credit responses are weighted using a formula that gives more importance to responses from institutions with greater worldwide exposure and more sophisticated country analysis systems." The creditworthiness ratings are preferable to credit ratings because they are available for many countries before they were formally rated. While other work has been able to extract "shadow ratings" for these countries, having a real measure of their perceived risk is helpful for detecting what will likely be subtle effects.⁴² On the spectrum of ratings, these ratings are especially perceptual because they are based on an average of individual surveys. I obtained the IIR ratings by scanning and transcribing the tables published in March and September issues of the U.S. edition, 1987-2012.

To study the behavior of democracy raters, I used the political rights score from Freedom House. In this way, I treat the Freedom House score as a perception of the country's level of democracy rather than as an objective measure of the true level of democracy. I flip the score so that higher levels refer to higher scores on political rights, from 1 to 7.

A table of summary statistics appears in the appendix.

Quantifying Classification Effects

Empirical Strategy

I model the relationship between classifications and outcomes using the following OLS regression:

⁴¹For precedent, see Büthe and Milner (2008).

⁴²For works comparing various measures of creditworthiness, see Vij (2005); Oetzel et al. (2001); Ratha et al. (2011).

$$Y_{i,t} = \alpha + \beta \text{Above cutoff}_{i,t-lag} + \delta \log(\text{GNIPC})_{i,t-lag} + \gamma \mathbf{X}_{i,t-lag} + \mu_i + \tau_t + \varepsilon \quad (1)$$

where Y is an outcome, \mathbf{X} represents a vector of covariates, t denotes the year or period of analysis, and lag refers to the lag between threshold-crossing and outcome measurement. Country and period/year fixed effects are used, and I cluster standard errors by country. This regression model is based on the specification used in similar work by Knack et al. (2014), who study the effect of crossing the operational cutoff on aid allocations.⁴³

My preferred specification of the model differs somewhat across the dependent variables. In keeping with Knack et al. (2014), I aggregate my observations into three-year periods to smooth volatile outcome variables, aid and FDI. In particular, I group logged observations into three-year periods corresponding with the IDA replenishment cycles of the World Bank.⁴⁴ This grouping therefore reflects the natural decision-making timeline of donors, who plan their allocations to the World Bank in three-year cycles. Because of the period grouping, the lag is just one period: a donor's decision in one replenishment cycle is based on the observed classification of the country in the last year of the last replenishment cycle. In the case of aid, I restrict the sample only to countries that have ever benefited from IDA during the time frame, since these are the only countries we would expect to experience change in their aid receipts. Since all countries may receive FDI, credit ratings, and democracy ratings, I use samples of all countries in these regressions.

Ratings data does not naturally exhibit the clumpiness that appears in aid and FDI data, so in these regressions, I am able to use the yearly data and improve my sample size. I determine the lags based on my understanding of the rating process used by each rater. Since both scores represent perceptions, they should respond more quickly to classification changes than would aid and FDI. In the case of the ratings of country creditworthiness from the *Institutional Investor*, these ratings are published in September and March. Since the income classifications are released on July 1, I average the subsequent September score with the March score from the following year, and lag the independent variables by just one year. In other words, the world finds out in 2015 that income data from 2014 was above a threshold, and this affects the average creditworthiness score from September 2015 and March 2016. I do not know whether Freedom

⁴³Despite using the same empirical model, the results I present in the next section will differ from those reported in Knack et al. (2014). This is because my sample includes two additional periods (data through 2015) that were not available at the time the authors conducted their analysis. A complete replication of the paper and more detailed explanation of our different findings appears in the appendix.

⁴⁴All variables are aggregated using means, while income variables and the cutoff dummy are taken from the final year in each period.

House arrives at its scores before or after July 1. Since a change in classification announced in July 2015 on the basis of 2014 income data could affect either the Freedom House score for 2015 or for 2016, I look at both lags of one and two years.

To control for factors that are widely thought to shape economic perceptions of countries, I control for lagged values of the Freedom House Political Rights score, log population, and logged gross capital formation.⁴⁵ In the case when the Freedom House score is the dependent variable, I omit this as a control.

Finally, to improve comparability across outcomes, I standardize all dependent variables.

Results

To illustrate the patterns across classifications and outcomes most effectively, I present my results as coefficient plots in Figure 2. The same results are also reported in Table 2.

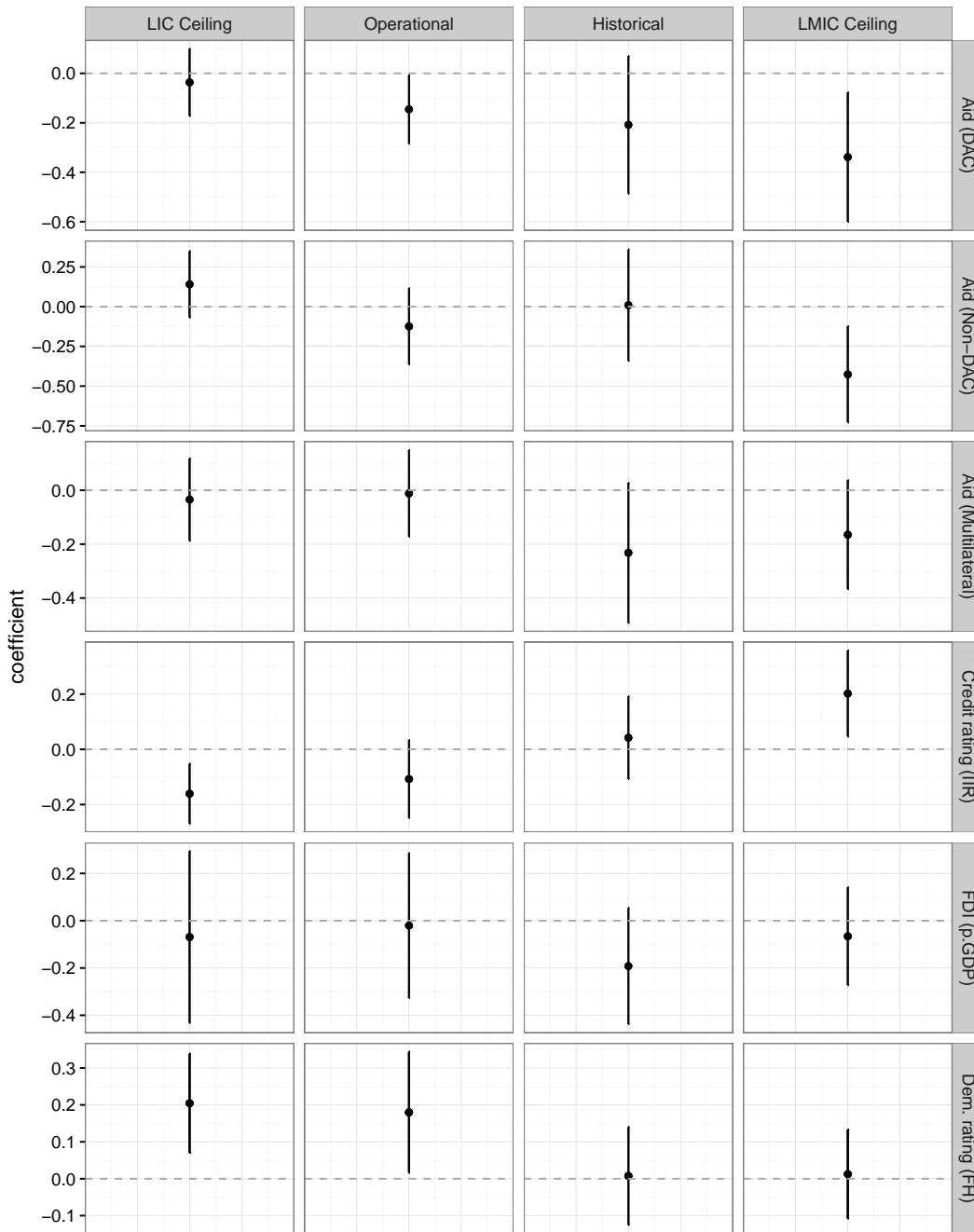
Consistent with the cognitive mechanism in hypothesis 1, I find that classifications affect raters. Crossing the LIC ceiling actually causes a country's creditworthiness rating to drop by about 2 points (non-standardized), while crossing the LMIC ceiling raises its rating by about 4 points.⁴⁶ Both results are significant at the .05 level. Crossing the LIC ceiling also inflates a country's democracy rating by about .4 points (non-standardized), and crossing the operational cutoff increases it by about .3 points. Both of these findings are most easily explained by the cognitive mechanism, as I predicted that heuristics would be visible in these perceptual measures. Furthermore, the effect on democracy ratings illustrates the "halo effect" at work: income should not even be an important consideration for assessing a country's respect for political rights. Income classifications appear not only to change perceptions of a country's income but also perceptions on unrelated attributes.

Consistent with the strategic mechanism in hypothesis 1, I also find that classifications affect donors. Crossing the operational threshold significantly decreases the amount of aid a country receives from traditional (DAC) donors by .16 standard deviations, but the effects of crossing the LMIC ceiling are more extreme. When this occurs, countries lose .32 standard deviations of aid from traditional donors and .52 standard deviations from non-traditional (non-DAC) donors, and both effects are statistically significant.

⁴⁵Gross capital formation is found by Vij (2005) to be one of the biggest predictors of the IIR rating.

⁴⁶I am currently investigating why crossing the LIC ceiling has a negative effect. While my theory predicts when and which classifications have an effect, it is agnostic about the direction. However, the negative result runs counter to my intuition and counter to the expectation of the Fellows I surveyed at the Center for Global Development prior to my analysis. A possible explanation is that crossing the LIC ceiling simply causes a country to be more likely to begin being rated by the IIR. Another explanation is that crossing the LIC ceiling raises expectations.

Figure 2: Coefficient plot of main results



Notes: Reported coefficients are from OLS regressions of the outcome (row) on a dummy variable coded 1 if a country is above the cutoff (column), controlling for GNI per capita. All regressions control for lagged values of log population, log gross capital formation, and Freedom House political rights score (Freedom House is omitted as a control when it becomes the dependent variable), and they include country and either year or period fixed effects. In the aid and FDI regressions, country-years are grouped into three-year country observations coinciding with IDA replenishment cycles; in ratings regressions, the unit of analysis is the country year. In the aid regressions, the sample is restricted to countries that have ever benefited from IDA after 1987. All dependent variables have been standardized for ease of comparison.

Table 2: Main regression results from preferred models

Outcome	Cutoff	Specification	Coef	SE	t	p	N	Lag	Cov:Pop	Cov:FHPR	Cov:gkform	Country	FE	Year	FE	Period	FE
ODA (DAC)	LIC Ceiling	Period (IDA-only)	-0.04	0.07	-0.53	0.60	631	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
ODA (DAC)	Op. cutoff	Period (IDA-only)	-0.15	0.07	-2.05	0.04	631	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
ODA (DAC)	His. cutoff	Period (IDA-only)	-0.21	0.14	-1.47	0.14	631	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
ODA (DAC)	LMIC Ceiling	Period (IDA-only)	-0.34	0.13	-2.54	0.01	631	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
ODA (Non-DAC)	LIC Ceiling	Period (IDA-only)	0.14	0.11	1.32	0.19	585	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
ODA (Non-DAC)	Op. cutoff	Period (IDA-only)	-0.12	0.12	-1.02	0.31	585	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
ODA (Non-DAC)	His. cutoff	Period (IDA-only)	0.01	0.18	0.05	0.96	585	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
ODA (Non-DAC)	LMIC Ceiling	Period (IDA-only)	-0.43	0.15	-2.76	0.01	585	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
ODA (Multilateral)	LIC Ceiling	Period (IDA-only)	-0.03	0.08	-0.45	0.66	629	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
ODA (Multilateral)	Op. cutoff	Period (IDA-only)	-0.01	0.08	-0.14	0.89	629	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
ODA (Multilateral)	His. cutoff	Period (IDA-only)	-0.23	0.13	-1.75	0.08	629	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
ODA (Multilateral)	LMIC Ceiling	Period (IDA-only)	-0.16	0.10	-1.60	0.11	629	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
Credit rating (IIR)	LIC Ceiling	Year (All)	-0.16	0.06	-2.91	0.00	2148	1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Credit rating (IIR)	Op. cutoff	Year (All)	-0.11	0.07	-1.49	0.14	2148	1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Credit rating (IIR)	His. cutoff	Year (All)	0.04	0.08	0.55	0.58	2148	1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Credit rating (IIR)	LMIC Ceiling	Year (All)	0.20	0.08	2.54	0.01	2148	1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
FDI (p.GDP)	LIC Ceiling	Period (All)	-0.07	0.19	-0.37	0.71	1064	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
FDI (p.GDP)	Op. cutoff	Period (All)	-0.02	0.16	-0.13	0.90	1064	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
FDI (p.GDP)	His. cutoff	Period (All)	-0.19	0.13	-1.53	0.13	1064	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
FDI (p.GDP)	LMIC Ceiling	Period (All)	-0.07	0.11	-0.62	0.53	1064	1	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
Dem. rating (FH)	LIC Ceiling	Year (All)	0.20	0.07	2.99	0.00	3056	2	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No
Dem. rating (FH)	Op. cutoff	Year (All)	0.18	0.08	2.16	0.03	3056	2	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No
Dem. rating (FH)	His. cutoff	Year (All)	0.01	0.07	0.11	0.91	3056	2	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No
Dem. rating (FH)	LMIC Ceiling	Year (All)	0.01	0.06	0.20	0.84	3056	2	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No

Notes: Reported coefficients are from OLS regressions of the outcome on a dummy variable coded 1 if a country is above the threshold and included covariates and fixed effects. All dependent variables have been standardized for ease of comparison.

While multilateral donors also appear to decrease aid, the result is just shy of conventional significance levels. These results are consistent with a strategic mechanism: donors care about being perceived as impartial by less informed audiences and therefore strategically allocate aid away from countries that have crossed into UMIC territory.

I observe no effect at all of country classification on the behavior of private investors. There is no statistically significant effect of any threshold on FDI flows. This is consistent with hypothesis 1 because these investors are less vulnerable to cognitive biases due to their need to justify investment decisions on the basis of profit potential, and they lack the need to appear impartial in their allocation decisions by using a third-party metric.

Of course, not every classification produces results. While I observe a few instances in which an operational threshold mattered, the results are stronger and more robust on the analytical thresholds. The LIC ceiling mattered especially for democracy ratings, and the LMIC ceiling had significant effects on creditworthiness ratings and aid allocations. This pattern is consistent with hypothesis 2, which suggested that analytical categories, which are simpler and more widely used, will be more influential.

In contrast, it is difficult to reconcile these results with the alternative explanation, which holds that classifications are powerful because they communicate information. The first implication of this explanation was that highly complex classifications would be more powerful, but this is not what I found. The income classification system is highly simplistic, and actually communicates even less information than would be gleaned by looking at the raw GNI data, which are publicly available. It is therefore surprising, by this explanation, that the LIC and LMIC ceilings would have effects. The second implication of this explanation was that actors at the greatest informational asymmetry would be most affected. Again, this is inconsistent with my findings. With field offices in the countries themselves, donors are highly expert actors when it comes to analyzing the state of development in a given country, but donors were one of the more responsive actors in my data. The informational explanation cannot account for this finding.

As described in the previous section, the empirical models used to estimate these coefficients differ slightly from one another to capture the unique timeline by which classifications would affect each actor's behavior. While this captures nuances in these relationships, one concern is that it limits comparability and permits cherry-picking significant results. In the appendix, I subject these results to several robustness checks by varying the sample, the period/year unit of analysis, included covariates and fixed effects, and different lags. As I discuss in the appendix, the results are remarkably robust to alternative specifications in

which we might logically observe a relationship.

Qualitative Evidence

Are these empirical findings consistent with the experiences of countries as they move from one category to another? To answer this question, in summer 2016, I interviewed a variety of stakeholders in the World Bank about proposed changes to policy regarding graduation from the World Bank's grant and concessional loans program, IDA. IDA graduation policy reform is currently one of the most pressing issues within the World Bank precisely because so many populous and geopolitically important countries are expected to soon graduate to Blend or IBRD status. Furthermore, my conversations occurred during an IDA replenishment year, so the topic provided a natural context for discussing these dynamics. Despite my focus on the operational categories, comments about the analytical income classification system also frequently arose. My 18 interviews included perspectives from the deputies donors send to IDA replenishment meetings, the executive directors of the World Bank or their advisors, country experts in the World Bank, and officials in IDA and IBRD. These individuals shared experiences representing or interacting with donor countries, recent graduates from IDA (now IBRD-only), blend countries under consideration for graduation, and IDA-only countries that have recently crossed the operational cutoff.⁴⁷

Several respondents confirmed my finding that aid losses from graduation extend beyond IDA. One of the current candidates for graduation, Vietnam, has already reported losing access to bilateral funds as a result of its candidacy.⁴⁸ The same has occurred for previous graduates Armenia and Georgia. An [official] stated that one donor backed out of a trust fund and cited Georgia's loss of IDA status as its reason for doing so. Furthermore, following Armenia's and Georgia's 2014 graduation, Georgia nearly lost its regional hub because it was viewed as less important to World Bank operations.⁴⁹ A World Bank Board official representing Armenia and Georgia noted that during 2016 negotiations over graduation policy reform, a proposal for access to an additional source of financing called IDA+ for recently graduated countries was withdrawn because "IDA is for the poorest countries."⁵⁰ Even the countries that have not yet become

⁴⁷My sampling protocol is available upon request. In conversations about both the views of donors and of recipients, I reached a point of saturation in which no additional information was learned. Because of this and because of my sampling protocol, I am able to generalize my findings to the dynamics involving all graduates within the last decade and all blend and IDA countries currently above the operational cutoff.

⁴⁸Author's interview with IDA deputy, July 22, 2016.

⁴⁹Citation temporarily removed.

⁵⁰Author's interview with advisor to Executive Director, August 3, 2016. The recent graduates were Angola, Armenia, Bosnia-

Table 3: Formal use of World Bank analytical classifications by other institutions

Organization	Use of World Bank classification
Global Fund to Fight AIDS, TB, and Malaria	Uses WB income data; uses LIC/LMIC threshold, splits LMICs into two groups at midpoint, uses LMIC/UMIC and UMIC/HIC thresholds to determine eligibility and concessionality ⁵⁴
Millennium Challenge Corporation	Uses WB income data; will not lend to any country over its own LMIC/UMIC threshold ⁵⁵
Gavi	Uses WB income data and own threshold (\$1,580 in 2016) to determine eligibility ⁵⁶
OECD DAC	All low and middle income countries as calculated by WB eligible; also all LDCs eligible; exceptions: G8, EU ⁵⁷

categorized as blend anticipate this dynamic. One advisor spoke of Zambian officials’ concern about losing access to aid, particularly after crossing the LMIC/UMIC threshold.⁵¹ In summary, one official stated, “Proposed IDA graduation reforms do not address the issue of perception of countries as IDA and non-IDA by the donor community.”⁵² The perception that donors coordinate on World Bank classifications was widespread; one IDA deputy stated, “we all know it goes on.”⁵³

Not only does this go on, but it is often institutionalized. Even though the World Bank never attached benefits to the analytical classifications, other multilateral development banks gradually began to use variants of them in their own policy decisions. In Table 3, I show that four significant sources of development finance explicitly incorporate the World Bank classification system into their eligibility policies.

This phenomenon of replicating World Bank income classifications in other organizations most strongly supports the strategic explanation because it represents the institutionalization of these classifications in policies. By using the classification scheme devised by a third party, these organizations are able to avoid some political backlash for their decisions. This behavior could not be explained by a cognitive bias: the process of formalizing eligibility requirements is highly contentious and could not be executed without careful and rational scrutiny.

My respondents’ experiences also conformed with my finding that classifications indeed affect beliefs by external actors about a country’s creditworthiness. An IDA officer formerly in charge of assessing the

Herzegovina, and Georgia.

⁵¹ Author’s interview with advisor to Executive Director, July 21, 2016.

⁵² Citation temporarily removed.

⁵³ Author’s interview with IDA deputy, July 22, 2016.

⁵⁴ http://www.theglobalfund.org/documents/core/eligibility/Core_EligibilityAndCounterpartFinancing_Policy_en/

⁵⁵ <https://www.mcc.gov/resources/doc/report-candidate-country-fy-2016>

⁵⁶ <http://www.gavi.org/support/apply/countries-eligible-for-support/>

⁵⁷ <http://www.oecd.org/dac/stats/daclist.htm>

creditworthiness of IDA graduates for IBRD lending noted that countries often ask for creditworthiness assessments as a “stamp of approval” to potentially help reduce their borrowing costs on the open market. This is because others know the Bank has a deep knowledge base and has done its homework.⁵⁸ Although the IBRD official suggested that creditworthiness assessments were fairly predictable, the process of assessing creditworthiness was much more likely to be described as a “black box” by stakeholders outside the IBRD, such as deputies, directors, and country desk staff, who often expressed confusion about why their countries were not found to be creditworthy. The opacity of this process, coupled with the perceived expertise of the World Bank, make a country’s eligibility to borrow from the IBRD an important signal to international observers.

In the case of the World Bank income classifications, countries’ responses were uniform: despite the benefits associated with being able to borrow from IBRD, countries nearly always resist graduation from IDA due to the steep losses in aid and borrowing volumes. However, my conversations suggest that countries may be more willing to sacrifice benefits in order to avoid categories such as the Fragile and Conflict Affected States and the Highly Indebted Poor Countries categories. The Kyrgyz Republic, as well as a few other states, have declined to take advantage of the HIPC program, despite their eligibility, because of a “political assessment that being classified as a Highly Indebted Poor Country was unfavorable.”⁵⁹ Further work should explore the conditions under which countries engage in this behavior.

Conclusion

Far from simply describing the world, classifications structure it in tangible ways. In this paper, I showed that when a country moves up in its World Bank income classification, donors give it less aid and credit and democracy raters improve its scores, but investors’ behavior is unaffected. Exploiting the exogeneity of the arbitrary thresholds that separate country categories, I demonstrated that these findings do not simply result from a relationship between income and economic outcomes, but are actually caused by the label a country receives from the World Bank. The patterns I observe in the data and my interviews offer strong support for my cognitive-strategic theory of classifications. For some audiences, such as the mass public and some ratings organizations, classifications act as a heuristic device to simplify a decision-making process. While institutional actors like donors and investors are less susceptible to these cognitive biases directly, they may

⁵⁸ Author’s interview with IBRD credit risk officer, July 26, 2016.

⁵⁹ Citation temporarily removed.

strategically use the classifications as a guide for allocation in an effort to demonstrate their impartiality to a less informed audience. My theory offers a new framework by which we can understand how a seemingly benign rhetorical tool can cause real behavior changes that materially affect the welfare of developing countries in the international economy. In so doing, I have illustrated that international organizations can exercise power through their bureaucracies, and are not merely the puppets of powerful state backers.

Documenting the existence of a “classification effect” raises important questions for future research. If classifications can affect the costs and benefits countries experience in the international economy, do countries strategically try to influence their classifications? Anecdotal accounts suggest they do, but future work can engage in a more systematic data collection effort to document and explain the various strategies and the conditions under which they are used. Also, if classifications are so powerful, how do international organizations wield this power, and to what ends? Although the World Bank income classifications appear to have acquired global influence unintentionally, it is plausible that other classification systems are engineered to incentivize certain behaviors. Studying the classifiers will provide a fuller account of this process.

My study also has implications for policy. When international organizations consider graduating or reverse graduating a country from a particular grouping or classification, they should take into consideration the expected response of external actors when estimating the impact of such a move. Moreover, international organizers should plan for the politicization of any classification they introduce, as my project illustrates that global development institutions, intentionally or unintentionally, affect how developing countries are perceived and treated in the international economy.

References

- Michael Barnett and Martha Finnemore. The Politics, Power, and Pathologies of International Organizations. *International Organization*, 53(4):699–732, 1999.
- Michael Barnett and Martha Finnemore. *Rules for the World*. Cornell University Press, 2004.
- Sarah M. Brooks, Raphael Cunha, and Layna Mosley. Categories, Creditworthiness, and Contagion: How Investors' Shortcuts Affect Sovereign Debt Markets. *International Studies Quarterly*, 59(3):587–601, 2015. ISSN 14682478. doi: 10.1111/isqu.12173.
- Tim Büthe and Helen Milner. The Politics of Foreign Direct Investment into Developing Countries: Increasing FDI through International Trade Agreements? *American Journal of Political Science*, 52(04):741 – 762, 2008.
- Allison Carnegie and Cyrus Samii. International Institutions and Political Liberalization: Evidence from the World Bank Loans Program. *British Journal of Political Science*, (Revise and resubmit), 2016.
- Michael A. Clemens and Michael Kremer. The New Role for the World Bank. *Journal of Economic Perspectives*, 30(1):53–76, 2016. ISSN 08953309. doi: <http://dx.doi.org/10.1257/jep.30.1.53>.
- Alexander Cooley and Jack Snyder, editors. *Ranking the World: Grading States as a Tool of Global Governance*. Cambridge University Press, 2015. ISBN 9780511973420. doi: 10.1029/2003JD004173.Aires.
- Allan Dafoe, Jonathan Renshon, and Paul Huth. Reputation and Status as Motives for War. *Annual Review of Political Science*, 17(1):371–393, 2014.
- Paul A. David. Clio and the Economics of QWERTY. *The American Economic Review*, 75(2), 1985.
- Axel Dreher, Jan Egbert Sturm, and James Raymond Vreeland. Development Aid and International Politics: Does Membership on the UN Security Council Influence World Bank decisions? *Journal of Development Economics*, 88(1):1–18, 2009. ISSN 03043878. doi: 10.1016/j.jdeveco.2008.02.003.
- Neil Fantom and Umar Serajuddin. The World Bank's Classification of Countries by Income. *World Bank Policy Research Working Paper*, 7528, 2016.
- M. Flinders and J. Buller. Depoliticisation: Principles, Tactics and Tools. *British Politics*, 1:293–318, 2006.
- Emmanuel Frot and Javier Santiso. Herding in Aid Allocation. *Kyklos*, 64(1):54–74, 2011.
- Julia Gray. *The Company States Keep: International Organizations and Investor Perceptions*. Cambridge University Press, 2013.
- Anthony G. Greenwald and Mahzarin R. Banaji. Implicit Social Cognition: Attitude, Self, and Stereotypes. *Psychological Review*, 102(1):4–27, 1995.
- Darren G. Hawkins, David A. Lake, Daniel L. Nielson, and Michael J Tierney. Delegation under Anarchy: States, International Organizations, and Principal-Agent Theory. In Darren G. Hawkins, David A. Lake, Daniel L. Nielson, and Michael J. Tierney, editors, *Delegation and Agency in International Organizations*. Cambridge University Press, Cambridge, 2006.
- Morten Jerven. *Poor Numbers: How We Are Misled by African Development Statistics*. Cornell University Press, Ithaca, NY, 2013.

- Daniel Kahneman. *Thinking Fast and Slow*. Farrar, Straus, and Giroux, 2011.
- Ashwin Kaja and Eric Werker. Corporate Governance at the World Bank and the Dilemma of Global Governance. *World Bank Economic Review*, 24(2):171–198, 2010. ISSN 02586770. doi: 10.1093/wber/lhq006.
- Judith G. Kelley and Beth A. Simmons. Politics by Number: Indicators as Social Pressure in International Relations. *American Journal of Political Science*, 59(1):55–70, 2015.
- Robert O. Keohane. *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton University Press, 1984.
- Erasmus K. Kersting and Christopher Kilby. With a Little Help from My Friends: Global Electioneering and World Bank Lending. *Journal of Development Economics*, 121:153–165, 2016. ISSN 03043878. doi: 10.1016/j.jdeveco.2016.03.010.
- Stephen Knack, Colin Xu, and Ben Zou. Interactions among Donors’ Aid Allocations: Evidence from an Exogenous World Bank Income Threshold. *World Bank Policy Research Working Paper*, 7039, 2014.
- Barbara Koremenos, Charles Lipson, and Duncan Snidal. The Rational Design of International Institutions. *International Organization*, 55(4):761–799, 2001.
- Ilyana Kuziemko and Eric Werker. How Much Is a Seat on the Security Council Worth? Foreign Aid and Bribery at the United Nations. *Journal of Political Economy*, 114(5):905–930, 2006. ISSN 0022-3808. doi: 10.1086/507155.
- Claudia Landwehr and Katharina Böhm. Delegation and Institutional Design in Health-care Rationing. *Governance*, 24(4):665–688, 2011.
- Bertin Martens. Why Do Aid Agencies Exist? *Development Policy Review*, 23(6):643–663, nov 2005. ISSN 09506764. doi: 10.1111/j.1467-7679.2005.00306.x.
- John J Mearsheimer. The False Promise of International Institutions. *International Security*, 19(3):5–49, 1994.
- Helen Milner. Why multilateralism? : foreign aid and domestic principal-agent problems. In Darren Hawkins, David Lake, Daniel Nielsen, and Michael J Tierney, editors, *Delegation and Agency in International Organizations*. Cambridge University Press, 2006.
- Scott Morris and Madeleine Gleave. The World Bank at 75. *Center for Global Development Policy Paper*, (58), 2015.
- Todd Moss and Benjamin Leo. IDA at 65: Heading Toward Retirement or a Fragile Lease on Life? *Center for Global Development Working Paper*, 246, 2011.
- Lynge Nielsen. How to Classify Countries Based on Their Level of Development: How it is Done and How it Could be Done. *IMF Working Paper*, 2011.
- Jennifer M. Oetzel, Richard A. Bettis, and Marc Zenner. Country Risk Measures: How Risky Are They? *Journal of World Business*, 36(2):128–145, 2001.
- Paul Pierson. Increasing Returns, Path Dependence, and the Study of Politics. *American Political Science Review*, 94(2):251–267, 2000.

- Dilip Ratha, Prabal K. De, and Sanket Mohapatra. Shadow Sovereign Ratings for Unrated Developing Countries. *World Development*, 39(3):295–307, 2011.
- Martin Ravallion. The World Bank: Why It Is Still Needed and Why It Still Disappoints. *Journal of Economic Perspectives*, 30(1):77–94, 2016. ISSN 0895-3309. doi: 10.1257/jep.30.1.77.
- James C. Scott. *Seeing Like a State*. Yale University Press, New Haven, 1998.
- Shelley E. Taylor. A Categorization Approach to Stereotyping. In David L. Hamilton, editor, *Cognitive Processes in Stereotyping and Intergroup Behavior*. Psychology Press, 1981.
- Madhu Vij. The Determinants of Country Risk Analysis: An Empirical Approach. *Journal of Management Research*, 5(1):20–31, 2005.
- Kenneth N. Waltz. *Theory of International Politics*. Waveland Press, 1979.
- R. Kent Weaver. The Politics of Blame Avoidance. *Journal of Public Policy*, 6(4):371–398, 1986.
- World Bank IDA Resource Mobilization Department. Review of IDA’s Graduation Policy. 2016.

A Summary Statistics

Table 4: Table of summary statistics

Group	Mean	SD	Min	Max	N
GNI per capita					
All	3619.85	8253.43	60.00	203900.00	3867
L	409.59	198.32	60.00	1020.00	1290
LM	1799.29	817.30	430.00	4050.00	1423
UM	5812.16	2388.77	1700.00	12550.00	878
H	21036.67	23462.91	6260.00	203900.00	276
ODA Disbursements (All donors)					
All	395.02	667.59	-959.96	11428.02	3846
L	651.80	863.64	11.84	11428.02	1290
LM	379.36	576.99	-943.15	5509.01	1423
UM	151.30	344.10	-959.96	3441.78	877
H	23.11	92.22	-460.26	559.30	256
ODA Disbursements (DAC donors)					
All	249.71	474.19	-967.54	10820.35	3846
L	399.28	652.72	2.22	10820.35	1290
LM	252.57	388.83	-967.54	2655.51	1423
UM	93.95	197.21	-455.12	1494.43	877
H	13.75	87.04	-458.17	557.03	256
ODA Disbursements (Non-DAC donors)					
All	13.33	119.41	-64.49	5166.11	3846
L	9.52	87.75	-64.49	2825.84	1290
LM	22.53	174.48	-38.37	5166.11	1423
UM	6.96	38.58	-22.05	836.01	877
H	3.19	20.60	-28.64	198.98	256
ODA Disbursements (Multilat. donors)					
All	131.99	238.55	-1171.55	3000.62	3846
L	243.01	274.77	-51.83	1982.71	1290
LM	104.26	193.96	-80.06	1780.36	1423
UM	50.40	216.33	-1171.55	3000.62	877
H	6.17	21.65	-56.40	186.87	256
FDI (percentage of GDP)					
All	3.95	7.31	-101.37	161.83	3703
L	3.32	6.91	-7.29	90.46	1238
LM	3.50	5.22	-32.29	62.30	1351
UM	4.98	5.47	-39.69	36.88	861
H	5.93	17.07	-101.37	161.83	253
Creditworthiness (HR)					
All	33.63	17.68	4.95	84.60	2370
L	20.47	9.48	4.95	63.10	791
LM	31.84	12.32	7.20	76.20	888
UM	47.03	14.90	14.85	82.35	556
H	67.30	11.95	23.40	84.60	135
Freedom House PR score (flipped)					
All	4.33	2.02	1.00	7.00	3723
L	3.22	1.68	1.00	7.00	1285
LM	4.50	1.86	1.00	7.00	1350
UM	5.31	1.87	1.00	7.00	858
H	5.83	2.08	1.00	7.00	230
Logged population					
All	15.49	2.10	9.19	21.03	3864
L	16.07	1.62	11.58	20.94	1287
LM	15.40	2.17	10.85	21.01	1423
UM	15.06	2.39	9.19	21.03	878
H	14.59	1.99	10.30	18.78	276
Logged gross capital formation					
All	21.46	2.91	0.00	29.20	3519
L	20.06	3.49	0.00	26.66	1206
LM	21.74	2.15	0.00	28.51	1234
UM	22.53	2.30	17.44	29.20	832
H	23.21	1.88	19.01	26.93	247

B Full Results

In the main paper, I reported results from the regressions that most logically modeled the relationship between a classification system and the specific actor it was influencing. One concern with this approach is that it limits the comparability of the results and encourages cherry-picking. In this section, I illustrate the robustness of my results to changes in the sample, unit of analysis, included covariates, and lags. Again, for ease of presentation, I illustrate the robustness checks using coefficient plots. In Figures 3-5, the black coefficients are the ones reported in the main paper, and the grey coefficients are from models in which one parameter is manipulated (the specific values of that parameter fall along the x-axis). The results are

reported in table form in Table 5.

In Figure 3, I vary the sample and unit of analysis. All else equal, I prefer using yearly data to improve my statistical power, however, I preferred grouping years into 3-year periods for the aid and FDI outcomes to smooth out data. Nonetheless, most of the aid and FDI results were very similar when analyzed using yearly data. My core finding – bilateral but not multilateral donors respond to the LMIC ceiling – held. None of the null results became significant except for the historical cutoff, which may cause multilateral donors to reduce their aid flows.

In parallel, I investigated what happened when I grouped my yearly rating observations into periods. My effects of the LIC and LMIC ceilings on credit ratings disappeared. The effect of the LIC ceiling on democracy rating remained robust, although the effect of the operational cutoff did not hold. In all, I still prefer the yearly data simply because it contains more information and there is no distributional need to smooth it.

I also examined the sensitivity of the results to changing the sample. All else equal, I prefer using data for all countries to avoid imposing artificial restrictions on the sample, but in the case of donor behavior, my preferred specification is to examine only the countries that have ever borrowed from IDA during the sample time period. This is because a donor’s decision to give aid to a country that has never borrowed from IDA during this time period is likely to be driven more by political considerations rather than considerations of recipient country need. Indeed, when I change the sample to all countries, some of the aid findings change: DAC donors are no longer responsive to the LMIC ceiling, non-DAC donors increase their aid to countries crossing the LIC ceiling, and multilateral donors actually increase the aid they give to countries crossing LMIC ceiling. The finding that non-DAC donors decrease their aid to countries crossing the LMIC ceiling is robust to the change in sample.

For the rating and investment outcomes, my preference is to include the full sample because all countries are expected to be rated and host investment, whereas only countries that have received IDA are really expected to receive aid. The puzzlingly negative effect of the LIC ceiling on creditworthiness again disappeared, but the positive effect of the LMIC ceiling on creditworthiness became even more pronounced when subset to IDA countries using yearly data. The positive effects of the LIC ceiling and the operational cutoff on democracy ratings also were robust to restricting the sample. The FDI results only changed when I restricted the sample and also switched to using yearly data, where in one instance, crossing the historical threshold significantly dampened FDI. However, I am skeptical of this result given the chunky distribution of the FDI data.

In Figure 4, I test the sensitivity of my results to dropping covariates and fixed effects. I would expect that some of my results may become null if the covariates were reducing the noise in the dependent variable that could be explained by observable characteristics. While imperfect, the results were surprisingly robust. The LMIC ceiling’s effect on DAC donors remained consistently negative, and its effect on non-DAC donors remained significant so long as fixed effects were included. The effects on the other outcomes also remained mostly robust to dropping covariates and fixed effects.

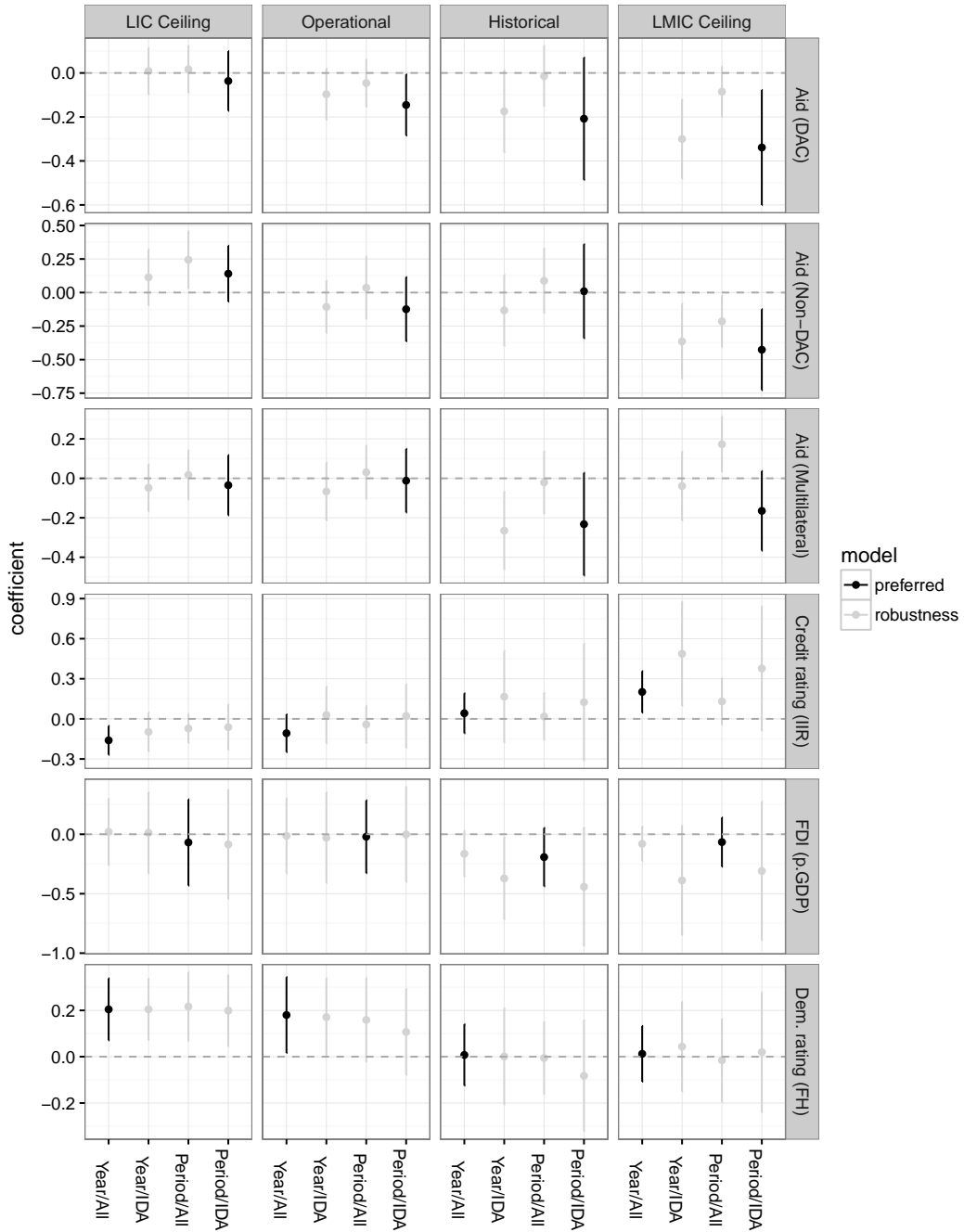
Finally, I test the sensitivity of my results on democracy ratings to changes in the lag in Figure 5. In general, I lag all period observations by 1, and I lag yearly observations by 2 for aid and FDI to account for time for decisions to change, and I lag yearly observations by 1 for the credit rating scores because the timeline of the IIR survey allows for perceptual changes to occur almost instantly. However, because I do not know the calendar used by Freedom House, I look at yearly lags of both 1 and 2 years. My findings were robust to either specification.

Table 5: All regression results from preferred models and robustness checks

Model	Outcome	Cutoff	Specification	Coef	SE	t	p	N	Lag	Cov:Pop	Cov:FHPR	Cov:gkform	Country FE	Year FE	Period FE
Preferred	ODA (DAC)	LIC Ceiling	Period (IDA-only)	-0.04	0.07	-0.53	0.60	631	1	Yes	Yes	Yes	Yes	No	Yes
Robustness	ODA (DAC)	LIC Ceiling	Period (All)	0.02	0.05	0.30	0.76	1073	1	Yes	Yes	Yes	Yes	No	Yes
Robustness	ODA (DAC)	LIC Ceiling	Year (IDA-only)	0.01	0.05	0.16	0.87	1764	2	Yes	Yes	Yes	Yes	Yes	No

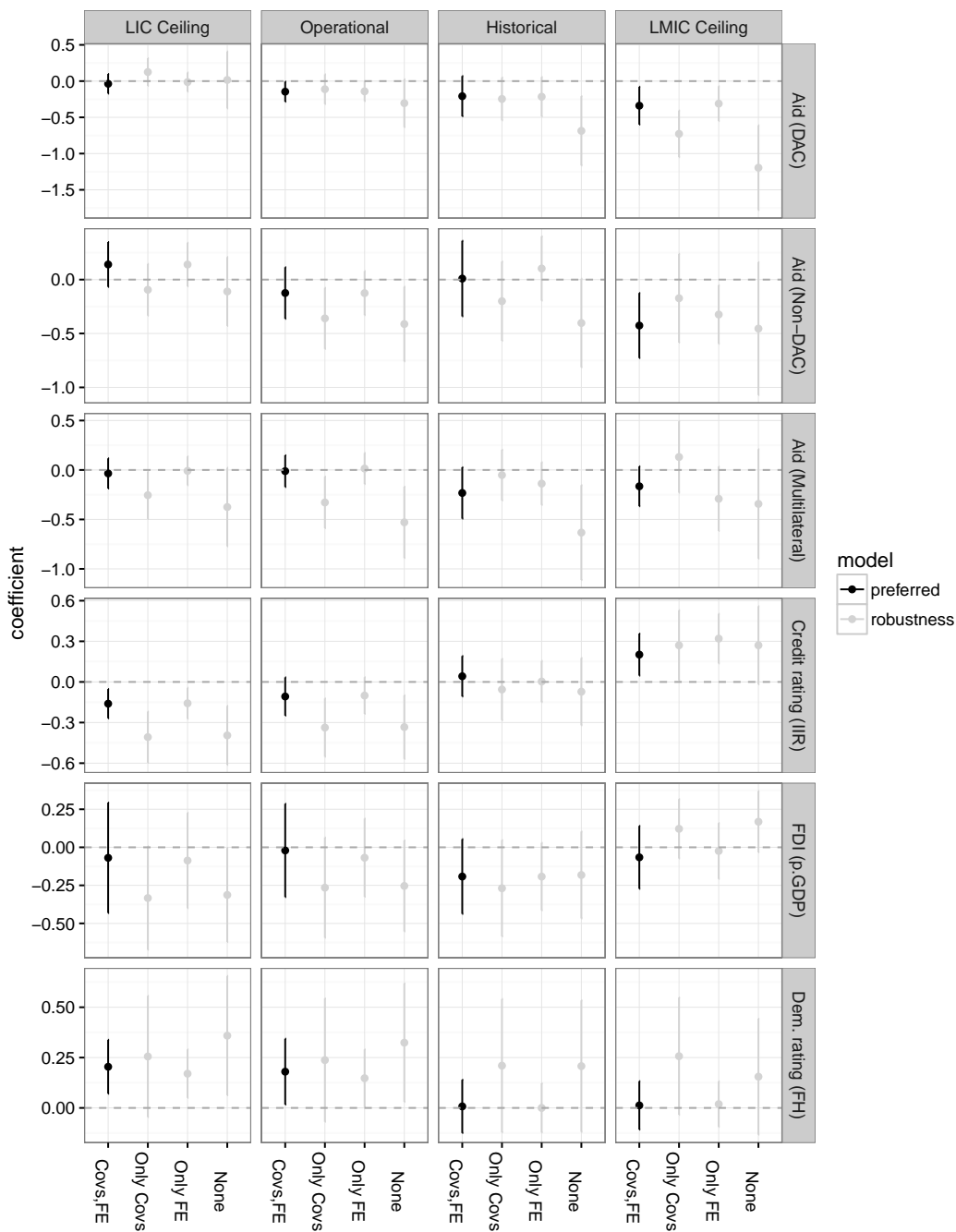
Robustness	Dem. rating (FH)	LMIC Ceiling	Lagged 1 unit	0.02	0.07	0.34	0.73	3204	1	Yes	No	Yes	Yes	Yes	No
------------	------------------	--------------	---------------	------	------	------	------	------	---	-----	----	-----	-----	-----	----

Figure 3: Robustness of results to changes in sample and unit of analysis



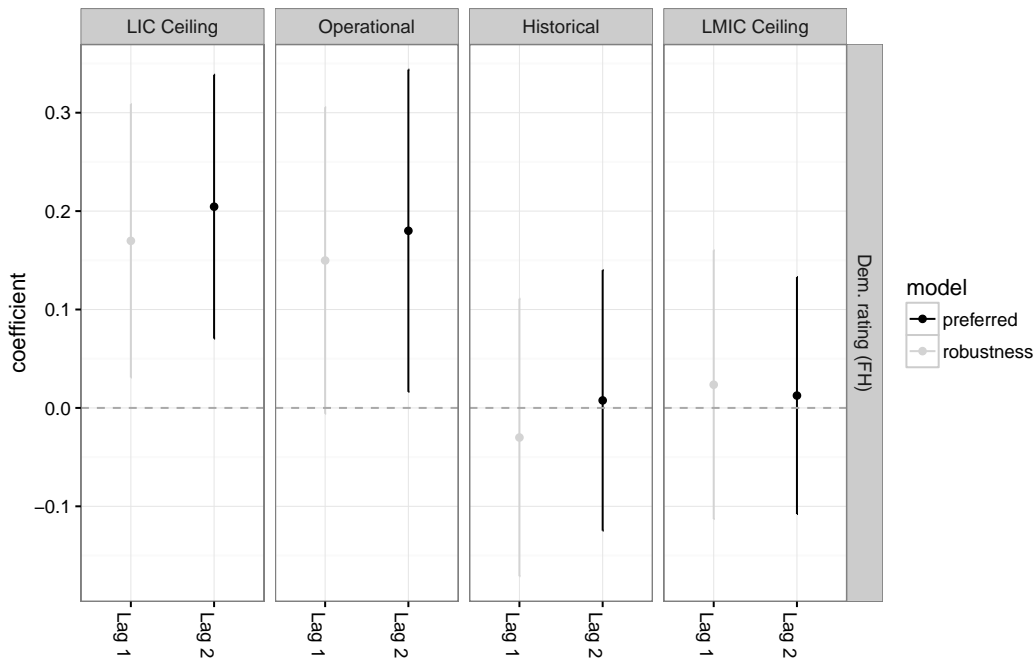
Notes: Reported coefficients are from OLS regressions of the outcome (row) on a dummy variable coded 1 if a country is above the cutoff (column), controlling for GNI per capita. All regressions control for lagged values of log population, log gross capital formation, and Freedom House political rights score (Freedom House is omitted as a control when it becomes the dependent variable), and they include country and either year or period fixed effects. All dependent variables have been standardized for ease of comparison.

Figure 4: Robustness of results to dropping covariates and fixed effects



Notes: Reported coefficients are from OLS regressions of the outcome (row) on a dummy variable coded 1 if a country is above the cutoff (column), controlling for GNI per capita. When included, covariates comprise lagged values of log population, log gross capital formation, and Freedom House political rights score (Freedom House is omitted as a control when it becomes the dependent variable). Fixed effects include country and either year or period fixed effects. In the aid and FDI regressions, country-years are grouped into three-year country observations coinciding with IDA replenishment cycles; in ratings regressions, the unit of analysis is the country year. In the aid regressions, the sample is restricted to countries that have ever benefited from IDA after 1987. All dependent variables have been standardized for ease of comparison.

Figure 5: Robustness of results to changing lag



Notes: Reported coefficients are from OLS regressions of the outcome (row) on a dummy variable coded 1 if a country is above the cutoff (column), controlling for GNI per capita. All regressions control for lagged values of log population and log gross capital formation, and they include country and year fixed effects. All dependent variables have been standardized for ease of comparison.

C Replication of Knack et al.

Knack et al. (2014) look at the effect of crossing the operational cutoff on aid disbursements from the OECD. They use data from 1987-2010 for all countries that either were IDA eligible or crossed the operational cutoff during this period, and group country-years into three-year periods that correspond with the IDA replenishment cycles. When aggregating to a period, they take the income and threshold observation for the final year in the period, and all other variables they take the mean of the three observations in the period.

There are three main differences between their model and data and mine:

1. They include Country Policy and Institutional Assessment (CPIA) index as a control. This is a score the World Bank awards to countries to evaluate economic policies, and this index is an important component in the formula that determines *allocation* of IDA, although not in eligibility decisions. While the CPIA's modern equivalent the IDA Resource Allocation Index (IRAI) is available publicly after 2006, the CPIA before 2006 is private data. Nonetheless, in correspondence with the authors, they confirmed that their results are robust to dropping this control.
2. The Knack et al. data end in 2010, which is the last year of the IDA15 replenishment cycle. My data end in 2015, which allows me to create observations for IDA16 and IDA17 as well.
3. There are minor differences in the observations included in the sample, mostly due to data availability. These observations are reported in Table 6.

In Table 7, I replicate the main results of the Knack et al. paper on the data provided to me by the authors and on my data. Model 1 is the benchmark result reported in Knack et al., although without the CPIA control. Even without this control, the result is significant: crossing the operational cutoff causes total

Table 6: Differences in included observations

In Knack et al. only	In Dolan only
Afghanistan—14	Bolivia—9
Afghanistan—15	Cameroon—9
Albania—10	Congo, Rep.—9
Angola—10	Honduras—9
Bhutan—9	Maldives—9
Bhutan—11	Papua New Guinea—9
Bhutan—12	Philippines—9
Bosnia and Herzegovina—12	Samoa—9
Congo, Dem. Rep.—9	Sao Tome and Principe—9
Congo, Dem. Rep.—10	Solomon Islands—9
Congo, Dem. Rep.—12	Tanzania—9
Congo, Dem. Rep.—13	Azerbaijan—10
Djibouti—11	Cameroon—10
Djibouti—12	Congo, Rep.—10
Dominica—9	Papua New Guinea—10
Dominica—10	Samoa—10
Dominica—11	Sao Tome and Principe—10
Equatorial Guinea—13	Turkmenistan—10
Equatorial Guinea—14	Ukraine—10
Eritrea—11	Haiti—11
Gambia, The—10	Papua New Guinea—11
Gambia, The—11	Sao Tome and Principe—11
Grenada—10	Syrian Arab Republic—11
Grenada—11	Ukraine—11
Kenya—9	Haiti—12
Liberia—12	Marshall Islands—12
Liberia—13	Micronesia, Fed. Sts.—12
Mongolia—10	Sao Tome and Principe—12
Montenegro—14	Haiti—13
Nicaragua—9	Marshall Islands—13
Nicaragua—12	Micronesia, Fed. Sts.—13
Nicaragua—13	Sao Tome and Principe—13
Serbia—13	Marshall Islands—14
Serbia—14	Micronesia, Fed. Sts.—14
Somalia—9	Sao Tome and Principe—14
South Africa—11	Marshall Islands—15
South Africa—12	Micronesia, Fed. Sts.—15
South Africa—13	
South Africa—14	
South Africa—15	
St. Kitts and Nevis—9	
St. Kitts and Nevis—10	
St. Kitts and Nevis—11	
St. Lucia—10	
St. Lucia—11	
St. Vincent and the Grenadines—9	
St. Vincent and the Grenadines—10	
St. Vincent and the Grenadines—11	
Sudan—10	
Sudan—11	
Turkmenistan—11	
Turkmenistan—12	
Turkmenistan—14	
Turkmenistan—15	
Vietnam—9	
Vietnam—10	
Yemen, Rep.—10	
Zimbabwe—13	
Zimbabwe—14	

ODA to decrease. When I run the same model on my own data set in Model 2, I obtain very similar results, even though I am missing observations that appear in the first column of Table 6. In Model 3, I expand the sample to include the observations that appear in the second column of Table 6. These are the observations that meet the criteria for inclusion in the Knack sample, but where not in the Knack data set. The results continue to hold.

However, in Model 4, I include two additional replenishment cycles for which data is now available. In practice, this extends the sample from 2010 to 2016. (I continue to include my observations from Model 2.) Adding more recent data causes the result to disappear entirely.

Given that Knack et al.'s results replicate perfectly but are highly sensitive to the inclusion of data published after their study, I also investigate some of their other findings, which are relevant for my argument. The authors also investigate the effects of the other thresholds that are included in my study: the LIC ceiling, the historical cutoff, and the LMIC ceiling. They do so by including these other dummies in the regression alongside the operational cutoff and find that each of these three other thresholds is not significant.

Table 7: Replication using alternative samples

DV: Log ODA from all donors				
	Thru IDA15	Thru IDA15	Thru IDA15	Thru IDA17
	(1)	(2)	(3)	(4)
Above op cutoff (0-1)	-0.21** (0.10)	-0.28*** (0.10)	-0.22** (0.10)	-0.13 (0.09)
Log GNI per capita (lagged)	-0.09 (0.14)	-0.04 (0.09)	-0.03 (0.09)	-0.17 (0.11)
Log pop (lagged)	-0.19 (0.40)	-0.04 (0.41)	0.14 (0.41)	0.09 (0.39)
Political rights (1-7, lagged)	0.04 (0.03)	0.05 (0.03)	0.04 (0.03)	0.05 (0.04)
Constant	25.43*** (6.98)	6.55 (6.02)	3.89 (5.99)	7.50 (6.73)
Observations	550	484	520	699
Country F.E.?	Yes	Yes	Yes	Yes
Period F.E.?	Yes	Yes	Yes	Yes
Data Source	Knack et al.	Dolan	Dolan	Dolan
Obs. included thru IDA15	Both+Knack	Both	Both+Dolan	Both+Dolan

Note:

*p<0.1; **p<0.05; ***p<0.01

In Table 8, I replicate their analysis on their sample (my data) and on my extended sample. None of the other thresholds have any significance in the sample ending with IDA15. However, the results change remarkably when we include observations through IDA17. Although the LIC ceiling remains insignificant, the historical cutoff approaches significance, and the LMIC ceiling is highly significant. These results hold regardless of whether the operational cutoff dummy is also included.

Evidently, how donors respond to these thresholds changes over time. One possible explanation for this is that the graduation process has become increasingly flexible in its application over time. In other words, there are more countries that cross the operational cutoff but do not change lending categories. If observers respond to lending categories, then the “treatment” of crossing the operational cutoff may attenuate. An alternative explanation is that other donors are less susceptible to the bias than they once were.

Table 8: Replication using other cutoffs

	DV: Log ODA from all donors								
	Thru IDA15	Thru IDA17	Thru IDA17	Thru IDA15	Thru IDA17	Thru IDA17	Thru IDA15	Thru IDA17	Thru IDA17
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Above op cutoff (0-1)	-0.26** (0.11)	-0.18 (0.12)		-0.29*** (0.10)	-0.13 (0.09)		-0.28*** (0.10)	-0.18** (0.08)	
Above LIC ceiling (0-1)	-0.05 (0.11)	0.08 (0.12)	-0.01 (0.08)						
Above his cutoff (0-1)				-0.14 (0.11)	-0.28 (0.18)	-0.28 (0.18)			
Above LMI ceiling (0-1)							0.02 (0.19)	-0.42*** (0.13)	-0.37*** (0.13)
Log GNI per capita (lagged)	-0.02 (0.09)	-0.19* (0.11)	-0.23** (0.11)	-0.001 (0.09)	-0.09 (0.13)	-0.14 (0.11)	-0.04 (0.09)	-0.10 (0.09)	-0.19** (0.08)
Log pop (lagged)	-0.05 (0.41)	0.08 (0.39)	0.13 (0.39)	-0.12 (0.40)	-0.09 (0.41)	-0.05 (0.40)	-0.04 (0.41)	-0.07 (0.37)	0.01 (0.37)
Political rights (1-7, lagged)	0.05* (0.03)	0.05 (0.04)	0.05 (0.04)	0.05 (0.03)	0.05 (0.04)	0.05 (0.04)	0.05 (0.03)	0.05 (0.04)	0.05 (0.04)
Constant	6.57 (6.01)	7.78 (6.79)	7.09 (6.78)	7.57 (5.94)	9.97 (6.94)	9.61 (6.92)	6.49 (6.05)	9.67 (6.28)	8.89 (6.38)
Observations	484	699	699	484	699	699	484	699	699
Country F.E.?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Period F.E.?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: *p<0.1; **p<0.05; ***p<0.01