Received: 14 August 2020 | Revised: 27 January 2021 | Accepted: 22 March 2021

DOI: 10.1111/gove.12605

ORIGINAL ARTICLE

Governance WILEY

Unimplementable by design? Understanding (non-)compliance with International Monetary Fund policy conditionality

Bernhard Reinsberg[1] | **Thomas Stubbs**[2] | **Alexander Kentikelenis**[3]

[1]University of Glasgow, Glasgow, UK

[2]Royal Holloway—University of London, Egham, UK

[3]Bocconi University, Milan, Italy

Correspondence
Bernhard Reinsberg, University of Glasgow, 40 Bute Gardens, Glasgow G12 8RT, UK.
Email: bernhard.reinsberg@glasgow.ac.uk

Abstract

Why do governments fail to implement policy commitments in contractual agreements with international organizations? While scholars have scrutinized domestic factors as obstacles to compliance, we argue that reform programs may be *unimplementable by design*. We study this hypothesis in the context of International Monetary Fund (IMF) programs, in which borrowing countries must commit to far-reaching economic policy reforms for access to credit. We collect detailed compliance data on individual policy conditions to assess the determinants of compliance failures of IMF programs from 1980 to 2009. Controlling for a host of borrower-specific variables, features of the loan, unexpected shocks during implementation, donor influence, and bureaucratic interest, we find that the number of conditions is a robust predictor of implementation failure. Our theoretical explanation for these findings is that over-ambitious program designs are the result of intra-organizational bargaining within the IMF bureaucracy. While an area department within the IMF drafts the initial reform program, functional departments use their amendment power to include policy conditions that they care about,

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2021 The Authors. *Governance* published by Wiley Periodicals LLC.

without due consideration of local circumstances, which leads to over-ambitious programs. These findings have important implications for theories of compliance as well as for policymaking in international organizations.

1 | INTRODUCTION

Lack of compliance with international commitments is widespread—not only for international agreements between governments, like the Paris Agreement or human rights conventions, but also for contractual agreements between international organizations and individual states (Chayes & Chayes, 1993; Findley et al., 2013; Hathaway, 2002; Tallberg, 2002; Underdal, 1998). One prominent example is implementation failure of borrowing countries with policy conditions of the International Monetary Fund (IMF), an international organization mandated to uphold global financial stability and provide loans to countries in economic trouble. The ability to tie disbursements to the fulfillment of conditions makes the IMF one of the most powerful international organizations (Stone, 2004; Vreeland, 2003); yet, its ability to compel actual policy reform in borrowing countries remains limited (Abbott et al., 2010; Haggard & Kaufman, 1992; Rickard & Caraway, 2019). Noncompliance with policy conditions is problematic for at least four reasons: first, it can undermine the borrowing country's ability to achieve macroeconomic stabilization (Nsouli et al., 2004); second, it can weaken market confidence in borrowing countries, thereby intensifying economic instability (Bird, 2002, pp. 838–839; Edwards, 2005; Mody & Saravia, 2006); third, it raises transaction costs by causing frequent re-negotiation of program content; and, fourth, it generates opportunities for donor influence and staff discretion in decision making, potentially thwarting program effectiveness (Dreher et al., 2015; Mecagni, 1999; Pop-Eleches, 2009; Stone, 2004).

Past scholarship examining the determinants of compliance failure has focused on country-specific factors, as well as unforeseen events in the global economy (Bird, 2002; Joyce, 2006; Stone, 2004). In contrast, we scrutinize the role of program design as an overlooked determinant of compliance. Introducing a new data set on compliance with IMF conditionality between 1980 and 2009, we argue that most IMF programs are *unimplementable by design*: they include too many and excessively complex conditions that cause compliance failures, in turn requiring the IMF to scale back on its ambitions in subsequent program reviews. Indicatively, of all 668 programs approved between 1980 and 2009, 371 programs were interrupted due to compliance problems, and 241 of these never resumed. In this article, we examine the determinants of permanent interruptions of IMF programs. Controlling for alternative explanations such as the initial macroeconomic environment, political circumstances, and the evolution of these factors, we find a significant positive relationship between program failure and the number of conditions set at the beginning of the program.

To explain these results, we draw on documentary evidence and semi-structured interviews to develop a theoretical framework that conceives the IMF as a collective agent within which several departments have a stake in program design. Any new program—designed by the relevant area department—can be amended by functional departments, which champion particular policies that they consider critical to the achievement of wider objectives. Functional departments can include "pet" issues in loan agreements without facing significant barriers because staff collectively prefer to avoid arbitration of disagreements by senior IMF leadership.

Meanwhile, IMF management and the Executive Board often do not monitor staff closely enough to reign in extensive conditionality. Borrowing countries accept over-ambitious programs to obtain much-needed IMF credit, while in many cases anticipating renegotiation or forgiveness on unmet conditions.

Our study contributes to scholarship on the policies and politics of international organizations, notably the IMF, which has received persistent attention in the literature. We demonstrate how entrenched bureaucratic structures inside the organization and the associated lending practices affect borrowing-country compliance. Our claim that IMF staff exercise major influence over design of conditionality conforms with constructivist works emphasizing the normative predispositions of staff, organizational culture, and the preference for ideological consistency (Barnett & Finnemore, 1999; Broome & Seabrooke, 2007; Chwieroth, 2014; Momani, 2007; Nelson, 2017). Yet, unlike these scholars, we shift attention to non-unitary rational actors involved in the intra-organizational bargaining process over the design of programs. In stressing such problems within a *singular* organizational structure, our argument is also distinct from those emphasizing collective action problems among *several* organizations participating in program design of bailouts (Chwieroth, 2014; Lütz et al., 2019). We therefore offer further support to the argument that organizational routines of international public bureaucracies affect the success of their policy interventions (Barnett & Finnemore, 1999). Finally, our study responds to calls for more reliable data and methods (Bird & Willett, 2004; Breen, 2014; Vreeland, 2006): we present a novel data set on program interruptions that we subsequently merge with publicly-available data on conditionality. Our data on interruptions extends existing sources—notably, the IMF's Monitoring of Fund Arrangements database—with respect both to the number of programs and time frame covered. We therefore advance earlier research restricted to proxies of compliance (Arpac et al., 2008; Dreher, 2003; Edwards, 2005) and provide convincing tests of our hypotheses. While we are not the first to use program interruptions as implementation measure, our large-N analysis covers more programs and a longer time horizon and offers methodological advances over past small-N research (Mecagni, 1999).

2 | DETERMINANTS OF IMF PROGRAM COMPLIANCE

While the design of IMF programs is commonly understood as the outcome of an unequal bargain between the Fund and its borrowers, compliance with such programs can be modeled as a decision-making problem by the borrowing government. In a simple cost–benefit framework, a government will comply as long as the marginal benefits exceed the marginal costs from doing so (Bird, 2008; Joyce, 2006; Stone, 2002). The standard explanation for non-compliance in this framework is that—as a program evolves—circumstances become less favorable than initially anticipated. An adverse shock may cause governments to fail to implement certain conditions even if they intended to comply at the point of agreeing to them (Bird, 2002).

One type of adverse shock relates to domestic *political* constraints, which often manifest in the form of anti-reform preferences of special interest groups (Coate & Morris, 2006; Drazen, 2002; Mayer & Mourmouras, 2005). Past research found that program interruptions are significantly related to the number of veto players (Arpac et al., 2008), and program disbursements often get interrupted before elections (Dreher, 2003). Another driver of non-compliance are *economic* shocks, such as a global financial crisis or declining terms of trade (Bird, 2008). To upset program implementation, shocks must not already be factored into program design. For instance, democratic governments often secure IMF programs with fewer conditions to begin with, given the anticipated domestic opposition to policy reforms (Caraway et al., 2012). This

suggests that the IMF anticipates potential implementation failures by designing programs that provide incentives for countries to comply with them.

In reality, however, asymmetric information may prevent IMF staff from identifying a program that respects these constraints (Mayer & Mourmouras, 2005). Non-compliance may occur where the net benefits of continuing an IMF program from a borrower perspective have become negative, while from the perspective of IMF staff they are still positive. Such situations may occur because borrowers do not internalize the positive externalities of crisis resolution beyond their borders. For example, while the IMF takes into account compliance-related benefits such as global financial stability beyond the program country, an individual country does not internalize these benefits but faces the full costs of adjustment (Joyce, 2006). In addition, the Fund does not fully anticipate political constraints. Our own interviews—in addition to internal evaluation reports and NGO briefs (BWP, 2018; IEO, 2018; IMF, 2018)—suggest that the Fund underestimates domestic political challenges to its conditions, with increased protests, government instability, and compliance failure as a result (Auvinen, 1996; Dreher & Gassebner, 2012; Ortiz & Béjar, 2013).[1] As a result, economic fundamentals and other characteristics of borrowing countries can affect program failure, even though program design is conditioned upon them.

Turning to empirical work on the determinants of program implementation, our analysis is closest to a landmark comparative case study of 36 Enhanced Structural Adjustment Facility programs in 1986–1994 (Mecagni, 1999). This study examined the reasons underlying all 51 identified cases of program interruptions: "policy slippages" (33 cases); government breakdown that made it impossible to continue the program (10 cases); and forward-looking disagreements with the Fund over reform schedules (8 cases). Further interrogating policy slippage cases, Mecagni (1999, p. 221) found that these were due to political events, economic shocks and natural disasters, but not program design and insufficient monitoring. The role of these factors was confirmed in subsequent studies—summarized in the Supporting Information (Appendix Table A1). While often using different proxies for program implementation, they also find that high debt, low exports, large loans, and political polarization *at program onset* adversely affect program implementation (Arpac et al., 2008; Joyce, 2006; Killick, 1995).

While previous scholarship offers a useful point of departure to understand variation in compliance with IMF programs, it has two shortcomings. First, it has understood non-compliance mostly due to changing circumstances outside the control of IMF staff and borrowing governments, and due to failure to accurately assess risks, thereby neglecting endogenous sources of program failure. Related empirical work has often relied on limited country samples, specific types of programs, and short timeframes, thereby limiting the generalizability of its findings. Second, previous scholarship on compliance conceives the Fund as a homogenous actor, thereby overlooking sources of non-compliance related to intra-organizational disagreements. In the next section, we address these shortcomings by presenting a new argument linking program design to compliance failure.

3 | A NEW HYPOTHESIS: UNIMPLEMENTABILITY BY DESIGN

While previous work has studied bargaining processes between the Fund and its borrowers over IMF programs, the internal processes affecting their design have been largely neglected. In fact, the process by which new IMF programs are agreed involves several stages and is relatively rigid. It begins with an area department preparing a blueprint, which is then circulated to the

functional departments, which can amend it. A revised blueprint is submitted to IMF management for clearance, after which an IMF mission enters into negotiations with the country. Any tentative agreement is again subject to approval by functional departments and ultimately the Executive Board, which triggers disbursement of the first credit tranche (Mussa & Savastano, 2000). A program then enters the implementation stage, where countries are subject to periodic review by IMF staff who assess whether the country has met all relevant conditions so that the next loan tranche can be released. If a borrowing country fails to meet a critical condition, the IMF Board can waive it so that the review is completed. Failure to complete the review leads to program interruptions and suspension of loan disbursements (see Supporting Information, Appendix A for a detailed description, and Box 1 for a glossary of key terms).

BOX 1 Glossary of key terms

Program design

Binding conditions: Policy conditions that a borrowing country must implement in the agreed time period. Binding conditions include prior actions, structural performance criteria, and quantitative performance criteria. Failure to implement a binding condition delays the program review and interrupts the program.

Nonbinding conditions: For these policy conditions, which include indicative targets and structural benchmarks, no waiver is required if they have not been met. Judging whether sufficient progress has been made on their implementation is at the discretion of IMF staff.

Program implementation

Program reviews: A monitoring mechanism whereby IMF staff periodically assess the progress of the borrower against agreed-upon reform measures. For a review to be completed, the borrower must be found to have implemented all binding conditions (and all unmet conditions must have been waived). The frequency of reviews differs by lending facility. For example, SBAs are reviewed quarterly, whereas EFFs are reviewed every 6 months. If the IMF staff cannot complete a review within the given review period, the respective program review is delayed and the program is interrupted.

Waivers: If a borrower fails to implement a binding condition, the IMF staff may recommend to the Executive Board to grant a waiver on that condition. The Board grants waivers on an ad hoc basis.

Program interruptions: A program is interrupted if its review is not completed by the test date. The schedule of test dates depends on the type of loan facility. There are two types of program interruptions, corresponding to two scenarios following a delayed review. A temporary interruption occurs if the country gets back on track, possibly after renegotiating the schedule of conditions. A permanent interruption occurs if the program does not resume. Both interruptions imply that the borrower loses access to IMF funding.

Sources: IMF, 2001a; Mecagni, 1999; Mussa & Savastano, 2000; Nsouli et al., 2004.

This short primer demonstrates that different IMF departments are involved in the process of program design. These departments have different preferences over program design, especially its overall level of ambition and the sets of issues included as policy conditions. The area department, represented by the mission chief, is the interlocutor with country authorities and thus has a better understanding of local circumstances. In contrast, functional departments—specifically, Strategy, Policy, and Review (SPR)—take a global view, often pushing for more ambitious targets and more conditions in IMF programs regardless of local circumstances. In an IEO survey on programs in fragile states, government representatives from Iraq attested that "the area department was more sympathetic to [their] situation than were the functional departments, which tended to be more rigid and to take an abstract view of what was achievable in Iraq with less understanding of local political context" (Takagi et al., 2018, p. 62).[2]

Given these preferences, conflicts between area departments and functional departments are inevitable. Asked about the internal coordination process on draft programs, a former mission chief explained: "There are disagreements about how fast adjustment can happen or about the need for conditions at all [...] SPR is usually harsher than the country unit." Such disagreements are likely resolved by stepping up the ambition of a program because "staff want to avoid involving IMF management—it is not good for their reputation."[3] Likewise, staff surveyed by IEO stated that "the internal incentive system rewarded toughness more than realism, and that negotiating a program with ambitious objectives and few departures from the mission brief smoothed the internal review process considerably, whereas attempts to be realistic and accommodative of [borrower] concerns—legitimate or not—did not" (IEO, 2002, p. 136). Thus, area departments will usually accommodate pressures from SPR for more ambitious programs.[4]

An observable implication of this process is that IMF programs are *unimplementable by design*: they will have extensive conditionality that even a well-intended government may not fully implement under favorable circumstances. This is obvious when considering that extensive conditionality is a result of the amendment power of functional departments. Narrative evidence that IMF programs have over-ambitious conditionality abounds (Baqir et al., 2005; Bird, 2005; Bird & Willett, 2004), although none of these studies analyzes the implications of this practice systematically. Referring to the 1994 program of the Philippines, an Independent Evaluation Office review stated: "[T]he IMF was simultaneously pushing for reforms to the oil-pricing system and to tax policy, each of which required congressional approval, as prior actions for the completion of program reviews. In the view of some staff, this may have been over-ambitious, exceeding the capacity of the political system to digest several major reforms at the same time" (IEO, 2002, p. 162).[5]

Even IMF management does not deny its programs suffer from unrealistic expectations. For example, a former staffer explained that "though the expansion of structural conditionality was a largely appropriate response to changing circumstances, there is a sense that we may have gone a bit too far," asking for "too much, too soon" (Dawson, 2003). Likewise, in its conditionality review, the IMF stated that "conditionality may have been established on policies that were unlikely to be delivered, calling into question the realism of program design" (IMF, 2001a).[6] It has also recognized the dangers of over-ambition for country compliance and socio-economic outcomes, given that "conditionality covering a broad range of policy areas may also place an enormous burden on limited administrative capacities in borrowing countries and may make it more difficult to focus on getting the most important things done" (Ahmed et al., 2001). The same study also warns "that conditionality that is unrealistically ambitious [...] may result in repeated failure to meet agreed targets and foster a culture of nonperformance." Thus, our main hypothesis is:

Hypothesis 1. The more conditions a borrowing country needs to implement in a given program, the greater the likelihood of implementation failure.

Why do governments sign onto over-ambitious programs? One explanation is that countries have limited influence on the agreement's terms, given they need credit to combat a crisis. However, even if they cannot influence these terms, they may find it rational to agree on over-ambitious terms. Countries are aware that programs are unrealistic and (rationally) anticipate implementation problems with some conditions, which the IMF may likely waive at a later stage (Pop-Eleches, 2009). A condition waiver is a decision by the Board to lift the requirement for the country to implement a given condition (see Box 1). This renders the net benefits from program participation positive again. The evidence is consistent with this argument: for instance, reflecting on a loan to Indonesia that contained 117 conditions, the IMF's official historian commented that "obviously, nobody expected Indonesia to fulfill all 117 of these promises. It was impossible, and everybody recognized it was impossible" (IMF, 2001b). The Indonesian case seems to reflect a general policy approach. Timothy Geithner—then-director of the predecessor of SPR—asked almost rhetorically: "Has the Fund been too tough or too accommodating? Or [...] has it been both, by setting unrealistic aspirations for policy reform and then acquiescing to the inevitable failure of even relatively well intentioned governments to meet the bar? There is something to this" (Goldstein et al., 2003, p. 442).

Why do IMF management and the Executive Board apparently fail to control over-ambitious program design by IMF staff? A straightforward explanation is they lack in-depth knowledge to judge whether programs are feasible. A less benign view is that in most cases they do not (sufficiently) care. A former mission chief explained that "management usually signs off without many questions asked—unless it is an important program." Once signed off by management, Board approval "is [...] perfunctory, [as it] will only care about big programs, like Greece," while rubber-stamping most staff proposals.[7] In general, institutional reforms seeking to curb the inflated use of conditionality among IMF staff, like the 2002 "Streamlining Initiative," were not particularly effective.[8] An IEO review stated that although the scope of conditionality narrowed, "there was little evidence of an actual reduction in the number of structural conditions" (IEO, 2004, p. 64); and more recent evidence shows the number of structural conditions remained high (Kentikelenis et al., 2016).

The final puzzle we address is why departments involved in program design do not internalize the costs of implementation failures. In fact, staff may suffer little reputational loss from designing over-ambitious programs that fail during implementation. A former mission chief confirmed that "the loss in reputation [from failure] is actually small," explaining further that "internally we do not play the blame game—unless a staff member is found guilty of serious misconduct."[9] One interviewee said "SPR is [like] Teflon."[10]

In sum, we hypothesize that extensive conditionality increases the likelihood of implementation failure, reflecting the notion that IMF programs are *unimplementable by design*. The subsequent section presents our research design.

4 | RESEARCH DESIGN

Our data cover 668 programs from 1980 to 2009 and is at the country-program level.[11] Our empirical approach is to use well-defined observational models in which we test for a relationship between program design and implementation failure, controlling for a host of alternative

explanations, such as macroeconomic fundamentals, unexpected economic shocks, and political dynamics.

4.1 | Main variables

4.1.1 | Implementation failure

While implementation failure can be measured in different ways, we argue that *program interruptions* are the most pertinent measure of implementation problems because they only occur if a borrower fails to implement critical policy conditions *and* the Fund decides not to waive them—that is, program interruptions occur when programs are seriously off-track. Among the measures discussed below and given data limitations, program interruptions are most suitable for our analysis.

To identify interruptions, we rely on a newly-compiled data set on IMF program review dates, which we used in conjunction with previously available data on IMF conditionality (Kentikelenis et al., 2016). We coded as interruptions those cases where scheduled reviews were not completed or completed with delay. A program review is delayed if there is a time lag between the initially agreed-upon review date and the actual review date, which vary by funding facility (Ivanova et al., 2001; Mecagni, 1999; Nsouli et al., 2004). For a Stand-By Agreement, a program review is delayed if it is not concluded within 90 days, while for an Extended Fund Facility, (Enhanced) Structural Adjustment Facility, or Poverty Reduction and Growth Facility program, it is delayed if not concluded within 180 days.[12] If a program review is never completed, the program becomes permanently interrupted. As there are more than one reviews in each program, a program may be both temporarily interrupted and permanently interrupted, although this co-occurrence is rare in practice.[13]

Our main dependent variable is PERMANENTLY INTERRUPTED—a binary indicator of whether an IMF program got permanently interrupted. This is the case if a scheduled review was never completed. For programs that are canceled and replaced with another, we did not count non-completed reviews as interruptions if the subsequent program commenced within 90 days of a Stand-By Agreement review date or 180 days of an Extended Fund Facility, (Enhanced) Structural Adjustment Facility, or Poverty Reduction and Growth Facility review date. In robustness tests, we use a more encompassing definition of program failure that also includes temporary interruptions.[14] Temporary interruptions occur if programs get off-track but ultimately resume after the Fund and the borrower have renegotiated the agreement. Our alternative outcome variable—ANY INTERRUPTION—thus takes a value of one if a program has at least one kind of interruption.

Program interruptions as dependent variable are preferable over several alternatives, listed in the Supporting Information (Table A2). Two measures of implementation failure are at the level of conditions. The MONA database records for each program condition an IMF staff assessment of whether it was implemented (Arpac et al., 2008). This information can be aggregated at the program level to obtain the number of *unmet conditions* as a measure of implementation failure. The main drawback of the MONA database is its short coverage, from 2002 to the present, which makes it inappropriate for our purpose. Another indicator of implementation deficits are *waivers*, which refers to decisions by the Executive Board releasing borrowers from the need to implement a condition that they failed to implement.[15] Waivers are less appropriate for our purposes as they are typically granted in the presence of mild implementation problems. Furthermore, the number of waivers is not a pertinent indicator of non-compliance because a

country might be non-compliant even without waivers, notably when its program is interrupted. Finally, waivers are better understood as an intervening variable—a tool to avert program failure once a borrower has failed to meet a critical condition. Given that waivers are an instrument to avert program failure, we include them as a control variable.[16]

A remaining alternative used by early studies on compliance is the *disbursement ratio*—the amount of funds drawn down by the recipient in relation to the originally committed funds—as a measure of compliance. A low disbursement ratio—typically below a threshold of 75%—may indicate non-implementation of essential targets. While Killick (1995) considers disbursement ratios of entire IMF programs, Dreher (2003) proposes a refinement that considers the percentage of funds left undrawn in any program year, assuming equal-size tranches. However, disbursement-based measures are too noisy proxies of compliance because countries may not need to draw on the funds; for example, if economic circumstances have improved. We therefore discard this measure for our analysis and use program interruptions hereafter.

In line with previous work, we find IMF program interruptions are surprisingly common (Mussa & Savastano, 2000, p. 94). Our data show that of all 668 programs during 1980 and 2009, 371 were interrupted, and 241 never resumed. In other words, 56% of all programs became interrupted over their lifetime, of which 28% had at least one temporary interruption and 36% were permanently interrupted, as shown in Table 1.[17]

4.1.2 | Program design

Our primary predictor relating to program design is the total NUMBER OF CONDITIONS. It captures the overall depth of conditionality and hence the ambition of the program (Dreher, 2009; Dreher & Jensen, 2007; Stone & Steinwand, 2008).[18] In robustness checks, we also consider the SCOPE OF CONDITIONALITY—the number of different issue areas in an adjustment program—as a proxy for program complexity (Stone, 2008). For all these measures, we only consider binding conditions, since implementation failure interrupts scheduled disbursements of IMF loans (Copelovitch, 2010). As conditions may be carried over to subsequent program years, we only consider conditions in the first year of the program, thereby mitigating concerns about reverse causality.

4.2 | Control variables

Considering both borrowing-country variables at program initiation and exogenous changes in the environment during program implementation, we probe five sets of control variables.[19]

TABLE 1 Program interruptions (1980–2009)

	Total number	Percentage
All programs	668	100%
Interrupted programs	371	56%
... Permanently interrupted	241	36%
... Temporarily interrupted	189	28%

Note: The set of programs excludes programs that cannot be interrupted, for example because they do not include conditionality. A program can have both types of interruptions.

First, in terms of economic variables, we include (logged) PER-CAPITA INCOME, GDP GROWTH, CURRENT ACCOUNT BALANCE, FOREIGN RESERVES, and DEBT SERVICE in percent of GNI (Arpac et al., 2008; Dreher, 2003; Pop-Eleches, 2009). The data for these variables are taken from the World Development Indicators. In addition, we control for changing economic circumstances in the global economy and the borrowing country, which may render initially agreed programs unviable (Bird, 2008). Considering the global financial environment, we include the percentage change in the U.S. INTEREST RATE given that re-financing becomes more difficult under higher U.S. rates. Data are from the Global Financial Development Database. At the country level, we include a dummy indicating a FINANCIAL CRISIS (Laeven & Valencia, 2013), as well as percentage changes in standard macroeconomic variables that the Fund closely monitors—CURRENT ACCOUNT BALANCE, RESERVES, and DEBT SERVICE as of GNI—calculated from the World Development Indicators.

Second, borrowing countries' domestic political characteristics may equally affect program design as well as implementation failure (Dreher, 2003; Joyce, 2006; Pop-Eleches, 2009; Stone, 2004). We include the POLITY IV index to measure democratic institutions (Marshall et al., 2010), the veto player index as a measure of POLITICAL CONSTRAINTS (Henisz, 2002), a binary indicator of LEFT-WING GOVERNMENT ideology (Beck et al., 2001), and an index of STATE CAPACITY (Hanson & Sigman, 2016). To capture changes in political characteristics, we include measures of the absolute difference in the POLITY IV score, the percentage change in POLITICAL CONSTRAINTS, dummies for EXECUTIVE ELECTIONS and LEGISLATIVE ELECTIONS (Beck et al., 2001), and the incidence of CIVIL WAR during program implementation (Gleditsch et al., 2002).

Third, we control for donor influence, given that salient borrowers for influential shareholder countries may obtain better deals while at the same time expending less effort to comply with conditionality (Breen, 2013; Copelovitch, 2010; Dreher & Jensen, 2007; Stone, 2002). Well-established measures of donor interest include a binary indicator of temporary UN SECURITY COUNCIL MEMBERSHIP (Dreher et al., 2015), UN GENERAL ASSEMBLY VOTE ALIGNMENT with the United States (Bailey et al., 2015),[20] and (logged) G5 BANK EXPOSURE (Copelovitch, 2010). Furthermore, we include (logged) ODA PER CAPITA to eliminate aid catalysis as a potential confounder, from the World Development Indicators. We measure all these variables at program initiation.

Fourth, we control for the IMF lending history of the borrowing country (and its peers), given the possibility for organizational learning from past failures.[21] We control for lending relations using the variables FIRST BORROWER—a binary indicator for the first program of a given country with the Fund—and PAST FAILURE—a binary indicator of any permanently interrupted program in the past 5 years. We also allow for the possibility that IMF staff learn from past failures from similar programs. We therefore include PAST PEER FAILURE, defined as the percentage of all permanently interrupted programs over the past 5 years that were managed by the same area department and that benefited countries in the same concessional bracket.

Finally, we control for additional dimensions of the IMF program, specifically the originally agreed PROGRAM DURATION in months and the LOAN TO QUOTA ratio (Arpac et al., 2008; Bird, 2008; Nsouli et al., 2004). Importantly, we control for the percentage of WAIVED CONDITIONS as waivers are intended to mitigate the risk of program failure. We obtained data on waivers by tracking the relevant Board decisions at all program reviews.[22] To account for global trends that apply to all borrowers alike, we include decennial period dummies. Finally, we include regional dummies to control for heterogeneity across area departments. Data sources and descriptive statistics for all variables can be found in the Supporting Information (Table A3).

4.3 | Methods

As our dependent variable is binary, we estimate probit models. The main drawback of such models is that they do not allow for the inclusion of country-fixed effects, known as the "incidental parameters" problem (Greene, 2011). Using fixed effects has the advantage of eliminating unobserved heterogeneity across countries, although the sample available for such analysis only includes countries with at least two programs. We use conditional logit models with country-fixed effects in robustness checks. In all cases, we compute robust standard errors clustered on countries.

We initially proceed under the assumption of exogeneity. This obviously overlooks the potential problem of reverse causality. In our case, however, reverse causality would work against our hypothesized relationship. For example, it might be the case that IMF staff reduce program ambition to avoid compliance failures—beyond a level that would be justified by observable measures of initial conditions and changing economic circumstances.

Another source of bias may arise from omitted variables that render our findings spurious. To use a classic example, "political will" is unaccounted for by the Fund when designing a program but affects country compliance (Vreeland, 2003). To mitigate such sources of bias, we adopt an instrumental-variable design. Following recent methodological advances (Stubbs et al., 2020), we use a "compound instrument" that identifies the differential effect of the number of conditions on program failure between borrowers likely to face many conditions versus those likely to face few conditions, controlling for covariates (Dreher & Langlotz, 2015). Specifically, we interact the number of countries under programs in the start year of a given program with the average number of program conditions for a given country over the sample period (Reinsberg et al., 2019). For the "excludability restriction" to be violated, there would need to be a variable predicting the number of program countries over time *and* differentially across countries with different general exposure to IMF conditions, which we consider to be highly unlikely. While diagnostic tests lend support to this claim (Figure B1), we also probe robustness of our findings to alternative instruments. Since we include the constituent effects of the compound instrument, we effectively control for global common shocks like a global financial crisis that might increase the number of IMF borrowers, as well as country-specific time-invariant effects related to the average burden of conditionality.

5 | RESULTS

5.1 | Correlational analysis

We find strong support for our hypothesis that over-ambitious program design leads to implementation failure (Table 2). Across all five models—corresponding to different sets of control variables—an increase in the number of conditions by one standard deviation is estimated to increase the likelihood of a permanent interruption by at least 2.4%. This effect is statistically significant ($p < .05$). While these findings are not to be interpreted causally, the robustness of the correlation is striking, providing a first indication that IMF programs are unimplementable by design.

A look at the control variables generates interesting insights on the remaining determinants of permanent interruptions. Condition waivers are negatively related to program failure. Other characteristics of the adjustment loan have no effect, except that programs benefiting low-

TABLE 2 Number of conditions and program failure

	(1)		(2)		(3)		(4)		(5)	
Permanently interrupted										
Number of conditions	0.013***	(0.005)	0.019***	(0.006)	0.012**	(0.005)	0.011**	(0.005)	0.012**	(0.005)
Waived conditions (%)	−0.022**	(0.010)	−0.040***	(0.013)	−0.032***	(0.009)	−0.023**	(0.011)	−0.022**	(0.010)
Program duration	0.006	(0.006)	0.010	(0.009)	0.007	(0.007)	0.007	(0.006)	0.006	(0.006)
Loan to quota (%)	0.000	(0.000)	0.001	(0.001)	−0.000	(0.000)	−0.000	(0.000)	0.000	(0.000)
Low-income country	0.206	(0.155)	0.451*	(0.255)	0.100	(0.195)	0.261	(0.166)	0.211	(0.155)
GDP per capita			0.170	(0.133)						
GDP growth			−0.026	(0.017)						
Current account balance			0.005	(0.009)						
Reserves			−0.108***	(0.039)						
Debt service (% GNI)			−0.006	(0.020)						
US interest rate (Δ) %			0.078	(0.118)						
Financial crisis			−0.384	(0.245)						
Current account balance (Δ) %			−0.012*	(0.007)						
Reserves (Δ)			−0.154**	(0.069)						
Debt service (Δ)			−0.001	(0.020)						
Polity IV					−0.021	(0.019)				
Political constraints					−0.270	(0.460)				
Left-wing government					0.042	(0.173)				
State capacity					0.043	(0.144)				
Polity IV (Δ)					0.012	(0.032)				
Political constraints (Δ)					−0.783	(0.512)				
Executive election					−0.155	(0.179)				
Legislative election					−0.260	(0.167)				

TABLE 2 (Continued)

	(1)	(2)	(3)	(4)	(5)
Civil war			−0.199	(0.277)				
UN Security Council member					−0.057	(0.202)		
UNGA vote alignment with United States					−0.043	(0.084)		
G5 bank exposure					0.017	(0.023)		
ODA per capita					−0.065	(0.072)		
First-time borrower							−0.096	(0.161)
Past failure							0.014	(0.125)
Past peer failure							0.000	(0.003)
Observations	599		387		528		592		599	
Pseudo-R^2	0.054		0.116		0.075		0.055		0.054	

*Significance level: $p < .1$. **Significance level: $p < .05$. ***Significance level: $p < .01$.

income economies have a marginally higher risk of failure. In terms of economic factors, low and declining levels of reserves and a deteriorating current account balance increase the likelihood of permanent interruptions. Neither domestic political factors, international political factors, nor bureaucratic factors are significantly related to program failure.

5.2 | Instrumental-variable design

Instrumental-variable analysis corroborates our main hypothesis (Table 3), producing coefficient estimates that are substantively larger and statistically significant. We find that the number of conditions robustly increases the likelihood of permanent interruptions. In substantive terms, based on the first model, an increase in the number of conditions by one standard deviation increases the likelihood of a permanent interruption by at least 11.5%—roughly half a standard deviation ($p < .05$).

Control variables change little compared to previous estimations. As the instrumental-variable design entails an additional equation for the number of conditions, it also produces interpretable findings on their determinants—itself an under-researched issue despite noteworthy exceptions (Copelovitch, 2010; Dreher et al., 2009; Dreher & Jensen, 2007). We find the compound instrument is moderately strong, with a Kleibergen-Paap *F*-statistic above ten (Stock & Yogo, 2005). We also find that longer programs tend to include more conditions already at program start, but a country that draws a larger portion of its quota (surprisingly) receives fewer conditions, which tends to be true also for low-income countries. In addition, the coefficients of economic fundamentals are consistent with expectations that programs include more conditions when the country must serve higher debt repayments. Turning to the political system, we find that countries where veto players mobilize and that suffer from civil war can expect lower conditionality. Among international political factors, higher G5 bank exposure (surprisingly) predicts more conditions, which could indicate that this variable proxies need for safeguards rather than political favoritism. Higher foreign aid available at the start of the program also predicts more conditions, for the same reason. Finally, first-time borrowers obtain significantly fewer conditions, while past permanent interruptions in the same region and income group tend to reduce program ambition.

5.3 | Mechanisms

Which mechanism underlies our finding that extensive conditionality leads to program failure? We argue that extensive conditionality is the result of the amendment power of IMF functional departments. Unfortunately, this mechanism is not directly testable in a large-N framework because there is no variation in the program design process. A potential remedy is to consider cases in which IMF policymaking departs from ordinary procedure.

To probe the plausibility of our collective-agency mechanism, we conduct split-sample analysis of the previous system of equations scrutinizing cases of elevated donor interest and IMF management oversight, where the relationship between program design and implementation failure would be weaker.[23] We consider two relatively simple proxies. The first is an annual measure of preference heterogeneity of the G7, which is computed as the coefficient of variation over the G7 ideal point estimates based on voting patterns in the UN General Assembly (Bailey et al., 2015). According to principal-agent theory (Nielson & Tierney, 2003), key shareholders

TABLE 3 Number of conditions and program failure taking endogeneity of conditions into account

	(1)		(2)		(3)		(4)		(5)	
Permanently interrupted										
Number of conditions	0.021**	(0.009)	0.036***	(0.012)	0.019*	(0.010)	0.020**	(0.010)	0.020**	(0.010)
Waived conditions (%)	−0.023**	(0.010)	−0.042***	(0.012)	−0.033***	(0.009)	−0.024**	(0.011)	−0.023**	(0.010)
Program duration	0.005	(0.006)	0.008	(0.008)	0.006	(0.007)	0.005	(0.006)	0.005	(0.006)
Loan to quota (%)	0.000	(0.000)	0.001*	(0.001)	−0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
Low-income country	0.224	(0.161)	0.522**	(0.257)	0.141	(0.209)	0.271	(0.171)	0.226	(0.162)
GDP per capita			0.202	(0.136)						
GDP growth			−0.024	(0.018)						
Current account balance			0.003	(0.008)						
Reserves			−0.110***	(0.037)						
Debt service (% GNI)			−0.017	(0.020)						
U.S. interest rate (Δ) %			0.055	(0.117)						
Financial crisis			−0.323	(0.243)						
Current account balance (Δ) %			−0.011	(0.007)						
Reserves (Δ)			−0.156**	(0.068)						
Debt service (Δ)			−0.008	(0.021)						
Polity IV					−0.022	(0.019)				
Political constraints					−0.223	(0.459)				
Left-wing government					0.036	(0.171)				
State capacity					0.061	(0.147)				
Polity IV (Δ)					0.013	(0.032)				
Political constraints (Δ)					−0.686	(0.514)				
Executive election					−0.172	(0.183)				
Legislative election					−0.246	(0.170)				

(Continues)

TABLE 3 (Continued)

	(1)		(2)		(3)		(4)		(5)	
Civil war					−0.170	(0.283)				
UN Security Council member							−0.049	(0.204)		
UNGA vote alignment with United States							−0.042	(0.084)		
G5 bank exposure							0.013	(0.023)		
ODA per capita							−0.063	(0.072)		
First-time borrower									−0.043	(0.175)
Past failure									−0.001	(0.127)
Past peer failure									−0.000	(0.003)
Number of conditions										
Compound instrument	0.020***	(0.006)	0.025***	(0.007)	0.026***	(0.006)	0.022***	(0.006)	0.025***	(0.006)
Countries under programs	−0.073	(0.149)	−0.115	(0.201)	−0.260	(0.171)	−0.132	(0.160)	−0.241	(0.147)
Condition intensity	−0.332	(0.366)	−0.639	(0.421)	−0.735*	(0.401)	−0.495	(0.386)	−0.657*	(0.350)
Waived conditions (%)	−0.024	(0.065)	0.028	(0.109)	0.052	(0.066)	−0.041	(0.067)	−0.032	(0.067)
Program duration	0.122**	(0.048)	0.077	(0.075)	0.133**	(0.053)	0.113**	(0.048)	0.115**	(0.046)
Loan to quota (%)	−0.007**	(0.003)	−0.007	(0.007)	−0.006*	(0.003)	−0.007**	(0.003)	−0.006*	(0.003)
Low-income country	−0.999*	(0.599)	−0.389	(1.257)	−1.668*	(0.936)	−0.193	(0.635)	−0.195	(0.645)
GDP per capita			−0.126	(0.908)						
GDP growth			0.064	(0.172)						
Current account balance			0.002	(0.050)						
Reserves			0.430	(0.287)						
Debt service (% GNI)			0.404**	(0.161)						
US interest rate (Δ) %			−0.166	(1.103)						
Financial crisis			−1.807	(1.409)						
Current account balance (Δ) %			−0.019	(0.076)						

TABLE 3 (Continued)

	(1)	(2)	(3)	(4)	(5)
Reserves (Δ)	0.554	(0.470)			
Debt service (Δ)	0.173	(0.160)			
Polity IV			0.091	(0.125)	
Political constraints			−1.696	(2.843)	
Left-wing government			0.809	(0.939)	
State capacity			−0.773	(0.881)	
Polity IV (Δ)			0.003	(0.191)	
Political constraints (Δ)			−9.143**	(4.522)	
Executive election			1.132	(1.249)	
Legislative election			−1.052	(1.139)	
Civil war			−2.567*	(1.388)	
UN Security Council member				0.405	(1.808)
UNGA vote alignment with United States				−0.244	(0.500)
G5 bank exposure				0.481***	(0.138)
ODA per capita				0.771*	(0.422)
First-time borrower					−6.657*** (1.246)
Past failure					1.204 (1.224)
Past peer failure					−0.049* (0.029)
Observations (Equation 1)	599	387	528	592	599
Pseudo-R^2 (Equation 1)	0.054	0.054	0.054	0.054	0.054
Observations (Equation 2)	599	387	528	592	599
R^2 (Equation 2)	0.439	0.441	0.426	0.438	0.475
F-statistic	11.102	11.928	16.740	12.425	19.811

*Significance level: $p < .1$. **Significance level: $p < .05$. ***Significance level: $p < .01$.

are less able to effectively control the IMF staff when their preferences are divided, which should give the IMF staff greater leeway to impose extensive conditionality.[24] We indeed find evidence of unimplementable IMF programs only when G7 preferences are relatively heterogeneous (Table A4). Our second test uses observable variation in the approval time of IMF programs. IMF management has discretion over which loans to schedule for approval by the Executive Board. It uses this discretion to accelerate loan approval for borrowers that it deems important (McDowell, 2017).[25] The observable implication is that our collective-agency mechanism should be driven by ordinary loans, with approval time above the median (Table A4). This is indeed what we find. Overall, these results provide some suggestive evidence for our collective-agency mechanism.

While our tests of the collective-agency failure mechanism are far from perfect, we do not find compelling evidence for alternative mechanisms. For example, extensive conditionality could reflect the preferences of powerful donors in the Executive Board. However, our own interviews, IMF documents, and previous research suggest that the Board is (more) lenient when it comes to program design (Bird & Willett, 2004). In a recent follow-up survey on the Streamlining Initiative, two-thirds of the Executive Directors said that conditionality could be more focused; and several of them expressed concern that the number of structural conditions and pace of structural reforms frequently overwhelmed country authorities, particularly in fragile states (IEO, 2018, p. 13). Other scholars emphasize that bureaucratic self-interest drives program design (Babb & Buira, 2005; Dreher & Vaubel, 2004; Vaubel, 2006). The IMF bureaucracy might be interested in mandating extensive conditionality because doing so provides it with increased discretionary authority to subsequently waive conditions (Eichenbaum, 2000). Prompted with this alternative, a staff member considered this a "far-fetched explanation."[26] Our interviewees also refuted that extensive conditionality is the result of "rational overshooting," whereby IMF staff impose more conditions than they expect borrowers to implement, whereas borrowers anticipate that IMF staff will waive unmet conditions. The fact that program interruptions are reputationally costly for recipients, for instance through increased refinancing costs, makes this argument unconvincing.

Overall, our results are consistent with our argument that extensive conditionality leads to implementation failure. Our core finding withstands different sets of control variables and estimation methods. Our posited relationship does not hold if donors have salient interests in the borrower and for loans that are important for IMF management. In both cases, political masters have high-powered incentives to rescue such borrowers, for instance through waivers of conditions that break the link between over-ambitious design and program failure.

5.4 | Robustness tests

In the Supporting Information, we probe the robustness of our findings in several ways. First, we use ANY INTERRUPTION as our dependent variable, thus defining program failure in broader terms than before (Table A5). Our results are qualitatively unchanged: a higher number of initial conditions is positively related to the likelihood of interruption. Control variables behave as before, except that an upcoming legislative election is now significantly negatively related to interruption.

Second, we use an alternative operationalization of our key predictor (Table A6). Specifically, we use the scope of conditionality, which reflects the complexity of a program in terms of the number of distinct policy areas in which the IMF requires reforms to be implemented

(Stone, 2008). Our core results are robust for both kinds of interruptions and across estimation methods: increased scope is significantly positively associated with program interruptions. Our positive findings on condition scope are consistent with the notion that over-ambitious design is a result of powerful functional departments that manage to include their favorite issues into programs.

Third, we control for nonbinding conditions, which may be related to the number of binding conditions while at the same time facilitating compliance because missing them does not automatically suspend the program (Table A7). We find that the inclusion of nonbinding conditions does not affect the relationship of interest for binding conditions. Whether nonbinding conditions directly affect program failure depends on the estimation method. If they reach statistical significance, their estimated effect is negative, which is consistent with the notion that nonbinding conditions have become increasingly popular with IMF staff to promote nonessential reforms in a flexible manner.[27]

Fourth, we use a conditional logit model with country-fixed effects. This is possible because many countries are repeat borrowers, with several IMF programs during the sample period, which allows us to control for unobserved confounders using fixed effects. In particular, fixed effects help eliminate the potential confounding effect of weak institutions as a joint determinant of extensive conditionality and program interruptions. In the models with endogeneity correction, we estimate a linearized outcome equation. Across various specifications, we obtain a robust positive relationship between the number of conditions and program failures (Table A8).

Fifth, we use two alternative instruments for the number of conditions. The first is the average number of conditions in all programs in the same region beginning in the same year. The intuition is that the IMF staff may compare countries under programs to similar peers and therefore prescribe a similar set of policies to countries within the same region (Broome & Seabrooke, 2007). A plausible micro-level process underpinning this argument is that the same staff are responsible for designing programs in the same region. The second instrument—in line with a world-polity approach—assumes that IMF staff follow global scripts in policy advice that they try to replicate in all countries alike (Kentikelenis & Seabrooke, 2017). We find both instruments to be strong predictors of the number of conditions, and results are qualitatively similar (Table A9).

Sixth, we verify that it is not the prevalence of *specific* kinds of policy conditions causing compliance failures. To this end, we sequentially include condition counts for issue areas such as public-sector reform, privatization of state-owned enterprises, price liberalization, external sector liberalization, financial sector reform, fiscal policy, institutional reform, and revenue measures (Table A10). Political economists and our IMF interviewees conceive most of these areas as controversial and therefore liable of causing program failure on their own (Haggard & Kaufman, 1992; Painter, 2005; Rickard & Caraway, 2019; Waterbury, 1992). While none of these affect our core result, we find that revenue measures reduce the risk of temporary interruptions, while price liberalization conditions and institutional reforms increase the likelihood of permanent failure.[28]

Finally, though suffering from reliability issues and short coverage, we replicate our main result using the number of unmet conditions as proxy for implementation failure, after aggregating relevant information for the conditions in a given program from the MONA data set. Data limitations restrict the analysis to the 2002–2009 period. Because the short time period does not allow us to run instrumental-variable regressions, we conduct two-way linear fixed effects regressions. We find a robustly positive association between the number of conditions at

program start and the total number of unmet conditions over the program lifetime (Table A11).[29]

6 | DISCUSSION AND POLICY IMPLICATIONS

The design of an IMF program is a coordination problem among IMF departments. While area departments design blueprints that try to accommodate local circumstances, the effective veto power of functional departments—especially, Strategy, Policy and Review (SPR)—leads to programs with ambitious targets and broad conditionality. Unless vital interests are at stake, both IMF leadership and the Executive Board rubber-stamp loan programs, while borrowing governments accept them given their immediate benefits in terms of fresh credit and the expectation of obtaining waivers for unmet conditions. As a result of this process, we posit that IMF programs contain extensive conditionality that makes them unimplementable by design.

We tested this hypothesis using a data set of 668 programs from 1980 to 2009 and established that the number of conditions in an IMF program is robustly positively related to the likelihood of program failure. Our findings withstand a battery of robustness tests and continue to hold when addressing potential endogeneity of conditionality. By emphasizing that most programs are unimplementable, our argument challenges existing views positing implementation as the norm, but not for unexpected negative shocks causing compliance problems that require modification of programs and granting of waivers (Joyce, 2006). While we find that IMF staff consider a wide array of observable characteristics—macroeconomic circumstances, domestic politics, global factors, and expected future realizations of all these variables—for initial program design, they fail to internalize the detrimental effects of their own intra-organizational bargaining, rendering programs unimplementable even for well-intentioned recipients.

While our explanation for over-ambitious program designs is inherently unobservable, our evidence is consistent with it in showing that the relationship between program design and implementation failure does not hold precisely when the vital interests of political masters are at stake, notably when borrowers are important to major donors or IMF management. In these cases, centralized oversight can remedy decentralized collective-agency failure. Yet, a weaker link between program design and implementation failure may also be bad news in the sense that borrowers anticipating lenient treatment become less faithful in implementing IMF conditions altogether, even when facing fewer conditions to begin with.

Before discussing avenues for future research and implications of our research, we note two limitations. First, some aspects of the compliance process remain unobserved. In particular, we do not know the exact set of conditions that governments failed to implement. Future research should take a disaggregated approach by examining variation in compliance with specific types of conditions. A disaggregated approach may be attractive for researchers willing to examine which sets of conditions are more demanding to borrowing countries in terms of implementation (Polak, 1991; Vreeland, 2006; Babb & Carruthers, 2008). Another promising avenue for further inquiry concerns the politics of waivers. While the presence of waivers implies that the government must have failed to implement the respective conditions, many unimplemented conditions remain unwaived. Under which conditions does the Fund decide to waive conditions? In this regard, it would also be important to study more closely the relationship between various compliance decisions. For instance, waivers—if granted exhaustively to cover non-

implemented conditions—can avert program interruption. But when does the Fund waive all conditions in full, and when does it only waive some conditions?

A second limitation is that our results identify the overall relationship between the depth of conditionality and program failure across all countries. Future research could take a case-based approach to process-trace the mechanisms underpinning over-ambitious program designs leading to implementation failure. While program interruptions are valid measures of implementation failure, they do not convey information about their causes. Although typically occurring as a result of failure to meet conditions, they can also be due to extraneous events, such as administrative delays or changes in political leadership.[30] Qualitative research could provide further insights into the mechanisms of program failure. With regard to temporary interruptions, there is also limited research into what factors help programs get back on track once they have been interrupted (Stone, 2004).

Future research should assess the relative weight of different explanations for extensive conditionality in IMF programs. We favor our collective-agency explanation because it makes few assumptions about preferences of relevant actors, interpreting extensive conditionality as the outcome of interactions among different departments with different incentives. While our preliminary tests provide support for the collective-agency mechanism, the evidence was weaker for other explanations, such as a hawkish Board, a general IMF staff preference for extensive conditionality, or the possibility of rational overshooting, whereby IMF staff impose more conditions than they expect recipients to implement and recipients accepting extensive conditionality knowing that IMF staff will waive unmet conditions. Yet, further research would be necessary to definitively pin down causal mechanisms.

Turning to wider implications, our research emphasizes policy design as a determinant of compliance overlooked by pertinent IR accounts of compliance. We posit a theoretical mechanism explaining why compliance is often poor, emphasizing internal fragmentation, coupled with rigid processes and misguided incentive structures within international organizations, rather than "lack of will" among recipient countries. Here, we draw on aid effectiveness research which has long noted that "donor fragmentation" may undermine effectiveness. To the best of our knowledge, no study has considered how *internal* fragmentation within international organizations undermines the effectiveness of their interventions.

Responding to calls to improve our understanding of the determinants of compliance with IMF policy prescriptions (Vreeland, 2006), our article uses new data and advanced methodology to explore this issue. Our results have implications for policymakers who aim to enhance the effectiveness of conditional lending assistance. If non-interruption is indeed crucial for program success (Mecagni, 1999), it would be important to reduce the incidence of program failure. A policy recommendation consistent with our analysis is to reduce the number of conditions and the complexity of programs. While it is difficult to say when there are "too many" conditions, reducing conditionality is beneficial for most countries, even though there may be cases where this strategy is ineffective. We therefore agree that "merely cutting back on the number of conditions would not in itself guarantee better implementation" (Mecagni, 1999, p. 238)—for that must be complemented with technical assistance and confidence-building measures for investors. The Fund should seize opportunities for *ex ante* modifying program design and program monitoring to help improve implementation (Mecagni, 1999, p. 238). Where over-ambitious program designs are the result of coordination failure among individual departments, member states need to strengthen the veto power of those parts in the IMF bureaucracy that articulate the collective interests of the Fund and its borrowing countries. Ultimately, organizational reform is an important first step to enhancing the effectiveness of IMF lending.

ACKNOWLEDGMENTS

We would like to thank Toke Aidt, Ursula Hackett, Chris Hanretty, Oliver Heath, Hyeran Jo, Andreas Kern, Melissa Lee, Ivica Petrikova, Cassilde Schwartz, Motoshi Suzuki, and participants at the 12th PEIO conference in Salzburg, the IBEI workshop on "IO conditionality" in Barcelona, and the PERG seminar at Central European University in Budapest for helpful comments.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

DATA AVAILABILITY STATEMENT

Replication data are available through Harvard Dataverse (https://doi.org/10.7910/DVN/QTFOGF).

ORCID

Bernhard Reinsberg https://orcid.org/0000-0001-7382-413X
Thomas Stubbs https://orcid.org/0000-0002-2624-9272
Alexander Kentikelenis https://orcid.org/0000-0002-1543-4595

ENDNOTES

[1] Interview #3. In responding to the 2018 IMF conditionality review, the Executive Board noted that IMF programs would underestimate the political difficulties related to implementation, further stressing "the need for more realistic implementation timetables and estimates of reform payoffs" (IMF, 2018, p. 3).

[2] Our interviewees confirm this, stating that "SPR is harsher than the country unit[s]" (Interview #1) and that "area departments are softer than SPR as IMF staff are going 'native'" (Interview #2). Philip Alston, UN Special Rapporteur on human rights and extreme poverty, argued "The Fund has consistently underestimated the importance of calibrating their recommendations to the specific political context, not taking into account the extent to which recommendations are politically viable and socially sustainable" (BWP, 2018, p. 2).

[3] Interview #1. Another interviewee reported that disagreements are not uncommon, occurring in about 25% of all programs (Interview #2).

[4] This is likely because SPR is more powerful, and the source of this power is its global knowledge and accumulated experience. As an interviewee confirmed: "SPR is the depository of knowledge; [the] area department may frequently defer to them" (Interview #2).

[5] One NGO interviewee attested that "over-optimism is embedded" (Interview #4), further citing how IMF chief economist Olivier Blanchard acknowledged that fiscal multipliers were wrong.

[6] More recent internal evaluations suggest that this problem has persisted, despite efforts to stem it. Responding to the results of the 2018 IMF conditionality review, Board members "saw scope for better tailoring and streamlining program objectives and structural conditions, particularly for fragile and small states, in light of their economic circumstances and capacity constraints" (IMF, 2018, p. 5). One NGO interviewee attested that IMF staff has become more aware about political difficulties to program implementation, albeit only in recent years (Interview #5).

[7] Interview #1. We acknowledge the possibility that the Board does not need to exercise tight control because IMF staff will take the preferences of important Board members into account when designing the program. Albeit not the focus of our analysis, we test this argument through our first-stage regressions on the determinants of program conditionality in the instrumental-variable setup and do not find significant associations of geopolitical variables and the number of program conditions.

[8] Indeed, as stated in an Executive Board response to the 2018 IMF conditionality review, such change "would require a change in culture, and continued adaptation and learning" (IMF, 2018, p. 5).

[9] Interview #1.

[10] Interview #2.

[11] The program is the observation unit but we consider dependencies across programs within a given country by clustering standard errors on countries.

[12] These are the typical review intervals used by the IMF. They are common knowledge to all parties and hence it is clear that if the borrower does not meet all binding conditions by the due date, the IMF staff cannot complete a review. Until a review is completed, it is delayed, and the next loan tranche cannot be released. Reviews therefore are important "test dates" over the program lifetime (IMF, 2001a).

[13] Fifty-nine out of 668 programs feature both kinds of interruptions. 371 programs feature any interruption.

[14] While programs may be temporarily interrupted more than once during their lifetime, we only consider the first interruption and record the year in which it occurred, discarding subsequent temporary interruptions. This serves to mitigate potential endogeneity bias.

[15] Waivers only exist for binding conditions. In contrast to nonbinding conditions, failure to implement binding conditions will suspend loan disbursements (IMF, 2001a).

[16] Despite waivers having the expected negative relationship with program failures, their inclusion does not affect the main results.

[17] These figures are consistent with previous research based on much shorter time series (Bird, 2002, p. 838; Ivanova et al., 2001, p. 7).

[18] Note that by counting the conditions we implicitly assign equal weight to them. This is unavoidable if we want to maintain replicability as attempts to code the substantive importance of conditions would introduce subjectivity bias.

[19] If the program was not interrupted, we measure changes between the year of program initiation and the year in which it was concluded. If it was interrupted, we consider changes between initiation year and the year in which the review that caused the interruption was scheduled.

[20] Results are similar for vote alignment with the G7 or including both alignment variables (Table A13).

[21] A 2005 IMF Managing Director report implies the theoretical possibility for such learning, arguing that "[p]arsimony must a guiding principle of conditionality for all client countries, not just those with a history of submitting to [the] IMF" (IMF, 2005, p. 6).

[22] A necessary limitation due to data availability is that counting waivers (as we do) abstracts from the burden a condition carries, given that waivers reflect not only non-implementation of a condition but also partial or delayed implementation.

[23] We prefer the split-sample approach over multiplicative interaction models as it avoids functional form misspecification and preserves the sanctity of the instrumental-variable design.

[24] The unconditional effect of preference heterogeneity is insignificant, which refutes the argument that extensive conditionality is driven by a divided principal itself.

[25] Ultimately, an important borrower may be one which is of geopolitical interest to major IMF donors, which makes shorter approval a fuzzy signal of management attention.

[26] Interview #1.

[27] Interviewee #1 highlighted the strategic value of nonbinding conditions, which do not need to be waived if not met, thus allowing programs to be kept alive.

[28] This is consistent with anecdotal evidence: price liberalization—as demanded by the IMF—is politically very costly, as the fall of regimes in the Arab Spring over the lifting of food price subsidies has shown (Interview #5). Our interviewees also mentioned political instability following elimination of fuel subsidies (Interview #4).

[29] We also coded program disbursements from the IMF web site. We find that the number of conditions is a robust predictor of whether a program has undrawn funds (Table A12), confirming our main results.

[30] Interviews suggest that most interruptions are "bad interruptions"—caused by bad design, bad policy decisions, and lack of commitment.

REFERENCES

Abbott, P., Andersen, T. B., & Tarp, F. (2010). IMF and economic reform in developing countries. *Quarterly Review of Economics and Finance*, *50*(1), 17–26.

Ahmed, M., Lane, T., & Schulze-Ghattas, M. (2001). Refocusing IMF conditionality. *Finance and Development*, *38*(4), 40–43.

Arpac, O., Bird, G., & Mandilaras, A. (2008). Stop interrupting: An empirical analysis of the implementation of IMF programs. *World Development*, *36*(9), 1493–1513.

Auvinen, J. Y. (1996). IMF intervention and political protest in the third world: A conventional wisdom refined. *Third World Quarterly*, *17*(3), 377–400.

Babb, S. L., & Buira, A. (2005). Mission creep, mission push and discretion: The case of IMF conditionality. In *The IMF and the World Bank at sixty* (pp. 59–84). London: Anthem Press.

Babb, S. L., & Carruthers, B. G. (2008). onditionality: forms, function, and history. *Annual Review of Law and Social Science*, *4*, 13–29.

Bailey, M. A., Strezhnev, A., & Voeten, E. (2015). Estimating dynamic state preferences from United Nations voting data. *Journal of Conflict Resolution*, *61*(2), 430–456.

Baqir, R., Ramcharan, R., & Sahay, R. (2005). IMF programs and growth: Is optimism defensible? *IMF Staff Papers*, *52*(2), 260–286.

Barnett, M. N., & Finnemore, M. (1999). The politics, power, and pathologies of international organizations. *International Organization*, *53*(4), 699–732.

Beck, T., Clarke, G., Groff, A., Keefer, P., & Walsh, P. (2001). New tools in comparative political economy: The database of political institutions. *World Bank Economic Review*, *15*(1), 165–176.

Bird, G. (2002). The completion rate of IMF Programmes: What we know, don't know and need to know. *The World Economy*, *25*(6), 833–847.

Bird, G. (2005). Overoptimism and the IMF. *The World Economy*, *28*(9), 1355–1373.

Bird, G. (2008). The implementation of IMF programs: A conceptual framework. *Review of International Organizations*, *3*(1), 41–64.

Bird, G., & Willett, T. D. (2004). IMF conditionality, implementation and the new political economy of ownership. *Comparative Economics Studies*, *46*(3), 423–450.

Breen, M. (2013). *The politics of IMF lending*. Springer.

Breen, M. (2014). IMF conditionality and the economic exposure of its shareholders. *European Journal of International Relations*, *20*(2), 416–436.

Broome, A., & Seabrooke, L. (2007). Seeing like the IMF: Institutional change in small open economies. *Review of International Political Economy*, *14*(4), 576–601.

BWP. (2018). *The political cost of IMF Programmes*. London: Bretton Woods Project. https://www.brettonwoodsproject.org/wp-content/uploads/2018/07/Observer_Summer_2018.pdf.

Caraway, T. L., Rickard, S. J., & Anner, M. S. (2012). International negotiations and domestic politics: The case of IMF labor market conditionality. *International Organization*, *66*(1), 27–61.

Chayes, A., & Chayes, A. H. (1993). On compliance. *International Organization*, *47*(02), 175–205.

Chwieroth, J. M. (2014). Professional ties that bind: How normative orientations shape IMF conditionality. *Review of International Political Economy*, *22*(4), 757–787.

Coate, S., & Morris, S. (2006). Policy conditionality. In S. Kosack, G. Ranis, & J. R. Vreeland (Eds.), *Globalization and the nation state: The impact of the IMF and the World Bank* (pp. 36–50). London and New York: Routledge.

Copelovitch, M. S. (2010). Master or servant? Common agency and the political economy of IMF lending. *International Studies Quarterly*, *54*(1), 49–77.

Dawson, T. C. (2003). *On the IMF's role and activities in Africa*. International Monetary Fund.

Drazen, A. (2002). Conditionality and ownership in IMF lending: A political economy approach. *IMF Staff Papers*, *49*(Special Issue), 36–67.

Dreher, A. (2003). The influence of elections on IMF Programme interruptions. *Journal of Development Studies*, *39*(6), 101–120.

Dreher, A. (2009). IMF conditionality: Theory and evidence. *Public Choice*, *141*(1–2), 233–267.

Dreher, A., & Gassebner, M. (2012). Do IMF and World Bank programs induce government crises? An empirical analysis. *International Organization*, *66*(02), 329–358.

Dreher, A., & Jensen, N. M. (2007). Independent actor or agent? An empirical analysis of the impact of U.S. interests on International Monetary Fund conditions. *Journal of Law and Economics*, *50*(1), 105–124.

Dreher, A., & Langlotz, S. (2015). Aid and Growth. New Evidence Using an Excludable Instrument. CESifo Working papers.

Dreher, A., Sturm, J.-E., & Vreeland, J. R. (2009). Global horse trading: IMF loans for votes in the United Nations security council. *European Economic Review*, *53*(7), 742–757.

Dreher, A., Sturm, J.-E., & Vreeland, J. R. (2015). Politics and IMF conditionality. *Journal of Conflict Resolution*, *59*(1), 120–148.

Dreher, A., & Vaubel, R. (2004). The causes and consequences of IMF conditionality. *Emerging Markets Finance and Trade*, *40*(3), 26–54.

Edwards, M. S. (2005). Investor responses to IMF program suspensions: Is noncompliance costly? *Social Science Quarterly*, *86*(4), 857–873.

Eichenbaum, M. (2000). Comment. In M. Mussa & M. Savastano (Eds.), *The IMF approach to economic stabilization* (pp. 122–128). NBER Macroeconomics.

Findley, M. G., Nielson, D. L., & Sharman, J. C. (2013). Using field experiments in international relations: A randomized study of anonymous incorporation. *International Organization*, *67*(04), 657–693.

Gleditsch, N. P., Wallensteen, P., Eriksson, M., Sollenberg, M., & Strand, H. (2002). Armed conflict 1946-2001: A new dataset. *Journal of Peace Research*, *39*(5), 615–637.

Goldstein, M., Geithner, T. F., Keating, P., & Park, Y. C. (2003). IMF structural programs. In M. Feldstein (Ed.), *Economic and financial crises in emerging market economies* (pp. 363–457). University of Chicago Press.

Greene, W. H. (2011). *Econometric analysis* (7th ed.). Pearson.

Haggard, S., & Kaufman, R. R. (1992). *The politics of economic adjustment: International constraints, distributive conflicts, and the state*. Princeton University Press.

Hanson, Jonathan K, and Rachel Sigman. 2016. *State capacity and World Bank Project success*. Ann Arbor: University of Michigan.

Hathaway, O. A. (2002). Do human rights treaties make a difference? *The Yale Law Journal*, *111*(8), 1935–2042.

Henisz, Witold J. 2002. *The Political Constraint Index (POLCON) Dataset*. Wharton University of Pennsylvania.

IEO. (2002). Evaluation Report: Evaluation of Prolonged Use of IMF Resources. Washington D.C.: Independent Evaluation Office of the International Monetary Fund.

IEO. (2004). *Evaluation of the poverty reduction strategy paper (PRSP) process and arrangements under the poverty reduction and growth facility (PRGF)*. Independent Evaluation Office of the International Monetary Fund.

IEO. (2018). *Structural conditionality in IMF-supported programs: Evaluation update 2018*. Independent Evaluation Office of the International Monetary Fund.

IMF. (2001a). *Conditionality in fund-supported programs*. Washington D.C.: International Monetary Fund.

IMF. (2001b). IMF conditionality: How much is enough? In *IMF Economic Forum*. International Monetary Fund.

IMF. (September 15, 2005). *The Managing Director's report on the fund's medium-term strategy*. Washington D.C.: International Monetary Fund.

IMF. (2018). *2018 review of program design and conditionality*. Washington D.C.: International Monetary Fund. https://www.imf.org/en/Publications/Policy-Papers/Issues/2019/05/20/2018-Review-of-Program-Design-and-Conditionality-46910.

Ivanova, Anna, Wolfgang Mayer, Alex Mourmouras, and George Anayiotos. 2001. *What determines the success or failure of fund-supported programs?*. Washington D.C.: International Monetary Fund.

Joyce, J. P. (2006). Promises made, promises broken: A model of IMF program Implementaion. *Economics and Politics*, *18*(3), 339–365.

Kentikelenis, A. E., & Seabrooke, L. (2017). The politics of the world polity: Script-writing in intergovernmental organizations. *American Sociological Review*, *82*(5), 1065–1092.

Kentikelenis, A. E., Stubbs, T. H., & King, L. P. (2016). IMF conditionality and development policy space, 1985–2014. *Review of International Political Economy*, *23*(4), 543–582.

Killick, T. (1995). *IMF Programmes in developing countries—Design and impact*. Routledge.

Laeven, L., & Valencia, F. (2013). Systemic banking crises database. *IMF Economic Review*, *61*(2), 225–270.

Lütz, S., Hilgers, S., & Schneider, S. (2019). When numbers don't add up: Bureaucratic culture and conflicts in EU-IMF programs. *Review of International Political Economy*, *26*(6), 1187–1210.

Marshall, M., Jaggers, K., & Gurr, T. R. (2010). *Polity IV project: Political regime characteristics and transitions 1800–2010: Polity level 2004–2008*. Center for Systemic Peace.

Mayer, W., & Mourmouras, A. (2005). The political economy of IMF conditionality: A common agency model. *Review of Development Economics*, *9*(4), 449–466.

McDowell, D. (2017). Need for speed: The lending responsiveness of the IMF. *Review of International Organizations*, *12*(1), 39–73.

Mecagni, M. (1999). The causes of program interruptions. In *Economic adjustment in low-income countries* (pp. 215–276). International Monetary Fund.

Mody, A., & Saravia, D. (2006). Catalysing private capital flows: Do IMF programmes work as commitment devices? *Economic Journal*, *116*(513), 843–867.

Momani, B. (2007). IMF staff: Missing link in fund reform proposals. *Review of International Organizations*, *2*(1), 39–57.

Mussa, M., & Savastano, M. (2000). The IMF approach to economic stabilization. *NBER Macroeconomics*, *14*(1), 79–122.

Nelson, S. C. (2017). *The currency of confidence: How economic beliefs shape the IMF's relationship with its borrowers*. Cornell University Press.

Nielson, D. L., & Tierney, M. J. (2003). Delegation to international organizations: Agency theory and World Bank environmental reform. *International Organization*, *57*(02), 241–276.

Nsouli, S. M., Atoyan, R., & Mourmouras, A. (2004). Institutions, program implementation and macroeconomic performance. *IMF Working Papers* WP/04/184.

Ortiz, D. G., & Béjar, S. (2013). Participation in IMF-sponsored economic programs and contentious collective action in Latin America, 1980-2007. *Conflict Management and Peace Science*, *30*(5), 492–515.

Painter, M. (2005). The politics of state sector reforms in Vietnam: Contested agendas and uncertain trajectories. *Journal of Development Studies*, *41*(2), 261–283.

Polak, J. (1991). The changing nature of IMF conditionality. *OECD Development Centre Working Papers*, Paris: OECD Publishing. https://doi.org/10.1787/187765075405.

Pop-Eleches, G. (2009). Public goods or political pandering: Evidence from IMF programs in Latin America and Eastern Europe. *International Studies Quarterly*, *53*(3), 787–816.

Reinsberg, B., Kentikelenis, A., Stubbs, T., & King, L. (2019). The world system and the hollowing out of state capacity: How structural adjustment programs affect bureaucratic quality in developing countries. *American Journal of Sociology*, *124*(4), 1222–1257.

Rickard, S. J., & Caraway, T. L. (2019). International demands for austerity: Examining the impact of the IMF on the public sector. *Review of International Organizations*, *14*, 35–57.

Stock, J. H., & Yogo, M. (2005). *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*. Cambridge University Press.

Stone, R. W. (2002). *Lending credibility: The International Monetary Fund and the post-communist transition*. Cambridge University Press.

Stone, R. W. (2004). The political economy of IMF lending in Africa. *American Political Science Review*, *98*(4), 577–592.

Stone, R. W. (2008). The scope of IMF conditionality. *International Organization*, *62*(3), 589–620.

Stone, R. W., & Steinwand, M. C. (2008). The International Monetary Fund: A review of the recent evidence. *Review of International Organizations*, *3*(2), 123–149.

Stubbs, T. H., Reinsberg, B., Kentikelenis, A. E., & King, L. P. (2020). How to evaluate the effects of IMF conditionality: An extension of quantitative approaches and an empirical application to government education expenditures. *Review of International Organizations*, *15*(1), 29–73.

Takagi, S., Donovan, D., Momani, B., Perez, L. L., de Las Casas, M., & Pisa, M. (2018). *The IMF and fragile states: Eight selected country cases*. Independent Evaluation Office of the International Monetary Fund.

Tallberg, J. (2002). Paths to compliance: Enforcement, management, and the European Union. *International Organization*, *56*(3), 609–643.

Underdal, A. (1998). Explaining compliance and defection: Three models. *European Journal of International Relations*, *4*(1), 5–30.

Vaubel, R. (2006). Principal-agent problems in international organizations. *The Review of International Organizations*, *1*(2), 125–138.

Vreeland, J. R. (2003). *The IMF and economic development*. Cambridge University Press.

Vreeland, J. R. (2006). IMF program compliance: Aggregate index versus policy specific research strategies. *Review of International Organizations*, *1*(4), 359–378.

Waterbury, J. (1992). The heart of the matter? Public Enterprise and the adjustment process. In S. Haggard & R. R. Kaufman (Eds.), *The politics of economic adjustment* (pp. 182–220). Princeton University Press.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Reinsberg, B., Stubbs, T., & Kentikelenis, A. (2022). Unimplementable by design? Understanding (non-)compliance with International Monetary Fund policy conditionality. *Governance*, *35*(3), 689–715. https://doi.org/10.1111/gove.12605