Individual Neutrality and Credible Commitment: A Generative Relational Theory of Self-Legislation

Wanhong HUANG
huangwanhong@serendip.ngo

Abstract

A state that undertakes to remain neutral is believed for reasons that have little to do with its sincerity. It will outlive the government that made the undertaking, its conduct is recorded and observed by many parties at once, and it holds sunk institutional investments that defection would destroy. An individual who undertakes the same thing has none of these. This paper asks how a private person, holding no office and backed by no sovereign, renders a commitment to non-capture credible to strangers. Three deficits are identified in the position of the isolated individual, concerning duration, observability, and the cost of exit, and the paper argues that the devices by which individuals address them all work in one way: they recruit other parties into the enforcement of a rule the individual states. Self-legislation, so understood, is legislation for a self and enforcement by a field. A trilemma follows. Among credibility, revisability, and retained authority over the determination of breach, a commitment device secures at most two: removing the capacity to defect is credible and unrevisable, declaring a rule one judges oneself is revisable and unbelieved, and delegating the judgment of breach is credible and revisable at the cost of the authority. The paper further distinguishes the registers through which trust is produced, treating symbolic guarantee, shared ordeal, and identification as distinct mechanisms with different failure conditions rather than as one process. Four instruments are compared: the international civil service, judicial recusal together with blind trusts and auditor independence, and a personal instrument that the author discloses and evaluates against the other three. The account is then tested against evidence that its own prescriptions fail, including the finding that disclosure of a conflict can increase rather than reduce the bias it reports. Two structural exposures are recorded: a short record makes an individual reputation both harder to establish and easier to destroy, and the community recruited to enforce a commitment is itself capturable.

Keywords: credible commitment; self-legislation; trust production; precommitment; private ordering; professional independence; neutrality.

Discussion Paper Note

This paper is a preliminary discussion paper intended to share an evolving idea and invite further dialogue, criticism, revision, and independent development. Its definitions, distinctions, and constructions remain provisional. Circulation across scholarly and practical communities is part of the purpose of releasing the manuscript at this stage.

The author treats the viewpoints, concepts, and lines of reasoning presented here as contributions to a shared field of inquiry. Similar or related ideas may have appeared in other intellectual, cultural, and disciplinary traditions. The manuscript therefore states its known antecedents, separates the researcher-origin proposal from later formal reconstruction, and leaves historical priority open pending a systematic originality review.

The arguments should be understood as provisional and historically situated. Readers are encouraged to question, test, revise, extend, reinterpret, or independently develop the ideas presented here. Where appropriate, acknowledgment of this paper as one point of encounter in the development of a related idea is appreciated. Such acknowledgment records an intellectual route; the ideas themselves remain available for criticism, revision, and independent development.

Responsible Use and Rights Reservation

This section separates requested scholarly conduct from the legal permissions stated on the following page. It records an ethical request for responsible use and then defines the narrower scope of retained legal rights.

The author encourages good-faith discussion, criticism, independent inquiry, and responsible use of the material in this work. Separately from the licence's terms, the author asks users to consider foreseeable harms when adapting or applying the proposed framework. This paper describes how commitments are made believable, and an account of that kind can be read as a manual for appearing trustworthy rather than for being so. The distinction is argued in the text and the author asks that it be preserved in any application. This ethical request leaves the licence's permissions and legally authorized uses unchanged.

The author retains the rights preserved under CC BY-NC 4.0 and may pursue remedies to which the author is legally entitled for breach of the licence or violation of the author's independently applicable rights. Reuse remains independent from authorial endorsement. Third-party rights require authorization from their respective holders where applicable. Copyright exceptions and limitations, including applicable forms of fair use or fair dealing, remain fully available.

Notices

This page consolidates the manuscript's publication status, licence, development disclosure, research-programme relation, declared interest, and suggested citation.

Status. This working draft records an evolving stage of the author's position and is circulated for discussion. Definitions, section structure, statements, and numbering remain subject to revision. Verification of several sources named in the accompanying literature audit, a systematic review of the philosophical literature on self-legislation, specialist review of the psychoanalytic material, empirical work on the demand-side claim, and an originality audit remain future research stages.

Licence. Except where otherwise indicated, copyright 2026 Wanhong Huang. This work is made available under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0). Subject to its terms, the licence permits sharing and adaptation for noncommercial purposes with appropriate attribution, a link to the licence, an indication of changes, and attribution that preserves the licensor's independence from the reuse. Reuse is governed solely by that licence; the responsible-use request on the preceding page remains separate from its terms. The licence deed and legal-code link are available at `https://creativecommons.org/licenses/by-nc/4.0/`. The licence governs in case of conflict with this summary. Third-party material remains subject to the rights held by its respective rights holders.

Statement on the use of language models. The exploratory discussions and preparation of this paper involved Anthropic's Claude. The model supported exploratory dialogue, source discovery followed by verification against publisher, journal, governmental, and institutional pages, argumentative criticism, and drafting in LaTeX. The author selected the research question, directed and approved the theoretical commitments and the epistemic status of the claims, and bears sole responsibility for the manuscript, including its definitions, constructions, taxonomy, arguments, conclusions, and errors. Authorship credit remains with the human author. The access level and claim limit for every cited source are recorded in the accompanying literature audit.

Declared interest. Section 8.3 examines an instrument of personal independence constructed by the author. The author therefore has an interest in the conclusions this paper reaches about instruments of that kind. The instrument is described in full, is evaluated by the same criteria applied to the other three cases, and is reported as failing several of them. The disclosure is made here because a paper arguing that commitments require verifiable disclosure would be self-undermining without it.

Related research programme. This paper is project P002 and the second paper in a series on trust, neutrality, and the transmission of shared experience. Project P001 takes responsibility for the account of neutrality as the governance of a field of relational conditions, for the constraint distinguishing the instruments a custodian may use from the influence it may exercise, and for the state-scale case. The present paper takes responsibility for the individual-scale credibility problem, the registers of trust production, the commitment trilemma, and the fragility analysis.

Suggested citation. Huang, Wanhong. "Individual Neutrality and Credible Commitment: A Generative Relational Theory of Self-Legislation." Working discussion paper, 2026.

1 Introduction

Consider two undertakings with the same content. A state declares that it will not align with either party to a conflict and will keep its territory, institutions, and offices available to both. A private individual declares the same. The declarations are identical in what they promise and unequal in what they are worth, and the inequality has nothing to do with the sincerity of either party.

The state will outlive the government that made the undertaking, and it cannot leave the system in which the undertaking was given. Its conduct is public, recorded, and observed continuously by many parties who compare what it does with what it said. And it holds investments that defection would destroy: hosting arrangements other states rely on, mandates other states have entrusted to it, a standing that took a century to accumulate and could not be rebuilt within a generation. None of this requires the state to be honest. It requires only that defection be expensive and visible.

The individual has none of the three. A person's undertaking expires with the person, and may be abandoned at any moment by simply going elsewhere. A person's conduct is observed by few and recorded by fewer, so that consistency maintained over years may leave no trace a stranger can consult. And a person who abandons an undertaking destroys nothing that anyone else was relying on, because a person who has made no arrangements that others depend on has provided nothing that could be withdrawn.

This paper asks what follows. If the credibility of a commitment rests on duration, observability, and the cost of exit, and if an individual lacks all three, then either individual commitments to non-capture are not credible, or they are made credible by some mechanism that supplies what the individual lacks. The paper argues for the second, and identifies the mechanism as recruitment: the devices by which individuals render commitments believable all work by drawing other parties into the enforcement of a rule the individual states. On this reading self-legislation is a misdescription of what occurs. The self supplies the rule; a field supplies the sanction; neither operates alone.

That conclusion generates the paper's central structural result. If the enforcement of a self-stated rule must be external, then the ways of arranging that externality are limited, and each arrangement purchases one property at the cost of another. Removing the capacity to defect is highly credible and leaves nothing to revise. Declaring a rule while retaining the judgment of one's own breach preserves revisability and persuades nobody. Delegating the judgment of breach to others is both credible and revisable, and transfers authority over one's own conduct to the parties one has recruited. The three properties of credibility, revisability, and retained authority are jointly unavailable, and any instrument of individual independence can be located by which two it has chosen.

The paper also distinguishes what is usually run together. Trust is produced by more than one mechanism, and the mechanisms have different requirements and different failure conditions. Trust may rest on a symbolic guarantee, in which a title, an office, a name, or a form of words is honoured because an order of signs stands behind it. It may rest on what a shared ordeal has revealed, where the parties have observed one another in circumstances that no account could substitute for. It may rest on identification, in which one party recognises itself in the other. These are not degrees of a single quantity, and an instrument that secures one supplies little of the others. A custodian who holds symbolic guarantee and nothing else is believed until the order that guaranteed it weakens; a custodian who holds only what an ordeal revealed cannot transfer it to a third party who was not present.

Four instruments are examined. The international civil service manufactures individual independence by placing it inside an office. Judicial recusal, the blind trust, and auditor independence manufacture it by removing a capacity or by requiring a disclosure. And the author has constructed a personal instrument of the kind this paper theorises, which is described, disclosed, and evaluated

against the same criteria as the other three, with its failures reported.

The account is written to be capable of failing and it partly does. The evidence that disclosure of a conflict can increase rather than reduce the bias it reports bears directly against the paper's own prescription that rules should generate verifiable artifacts. The evidence that communities of traders, neighbours, and criminals sustain credible commitments without any sovereign bears against the premise that the individual position is uniquely disadvantaged. Both are treated where they arise rather than in a closing concession, and both narrow the account rather than leaving it intact.

Section 2 supplies the regimes and the vocabulary the argument presupposes. Section 3 locates the account among the literatures on trust production, costly signalling, precommitment, self-legislation, diachronic agency, enforcement without a sovereign, role morality, and relational ontology, and states what each already establishes. Section 4 states the method and the conditions of disconfirmation. Section 5 develops the credibility problem. Section 6 distinguishes the registers of trust production. Section 7 develops self-legislation, relational enforcement, and the trilemma. Section 8 compares the four instruments. Section 9 treats fragility, entrapment, and depletion. Section 10 states what the analysis returns to the wider framework, Section 11 records the limits, and Section 12 consolidates the position.

2 Background and Preliminaries

This section supplies the material the argument presupposes. It describes the regimes through which individual independence is currently manufactured, states what an office holds that a person does not, and defines the vocabulary carried forward from the preceding paper. Analysis of these materials is reserved for Section 5 and Section 8.

2.1 Existing Regimes of Individual Independence

Individual independence is already manufactured, in several domains, by established means. Four are relevant here.

The international civil service supplies the most developed instance. Article 100 of the Charter of the United Nations provides that in the performance of their duties the Secretary-General and the staff shall not seek or receive instructions from any government or from any other authority external to the Organization, and shall refrain from any action which might reflect on their position as international officials responsible only to the Organization; the second paragraph binds each member state to respect the exclusively international character of those responsibilities and to refrain from seeking to influence the staff in discharging them [62]. The obligation is given individual form in the Staff Regulations, under which staff members are international civil servants whose responsibilities are exclusively international rather than national, and under which each staff member makes a written declaration, witnessed by the Secretary-General or an authorized representative, promising to exercise the entrusted functions in loyalty, discretion and conscience, to regulate their conduct with the interests of the Organization only in view, and neither to seek nor to accept instructions from any government or other external source [63]. The Regulations further provide that personal views and convictions remain inviolable while requiring that they not adversely affect official duties [63]. These provisions are elaborated for staff in a Secretary-General's bulletin on status, basic rights and duties [64].

Three features of this regime are recorded here and used later. The undertaking takes the form of a personal declaration. The declaration is witnessed rather than merely published. And the

declarant's independence is secured not only by the declaration but by a reciprocal obligation on the parties who might otherwise seek to capture them.

Judicial recusal supplies a second instance, in which independence is secured by removing the officeholder from the particular matter rather than by any undertaking about future conduct. The blind trust supplies a third, in which the officeholder's capacity to know their own holdings is removed so that the holdings cannot influence decisions; the device attracts sustained criticism on the ground that the officeholder knows what was placed into the trust and that the arrangement therefore functions as an appearance rather than a constraint [12]. Auditor independence supplies a fourth, framed in professional standards as independence in fact together with independence in appearance.

2.2 Capacities Held by an Office and Capacities Held by a Person

An office and a person differ in three respects that bear on whether an undertaking either makes is believed.

An office persists beyond the tenure of any occupant, and it cannot withdraw from the system in which its undertaking was given. A person's undertaking expires with the person and may be abandoned by departure. The relevant property is duration, and it belongs to the office rather than to the undertaking.

An office acts on a record. Its decisions are documented, its conduct is observed by many parties at once, and the parties compare what it does with what it said. A person's consistency, however long maintained, may leave no trace that a stranger can consult. The relevant property is observability.

An office holds arrangements that others rely on and that defection would destroy. A person who has made no arrangements on which others rely has nothing whose withdrawal would cost anyone anything. The relevant property is the cost of exit.

Luhmann's distinction between personal trust and system trust names the consequence [41, 42]. Trust in an office is trust in a system, held without the truster reconsidering it at each encounter; trust in a person is a decision taken in awareness of the alternatives, and taken again. Section 5 argues that the three properties are held by the isolated individual and not by the individual as such, and Section 7 argues that a community can supply all three.

2.3 Vocabulary Carried from the Preceding Paper

The account uses a small vocabulary established in the companion paper and restated here at the length required to follow the argument.

A *relation* is an ongoing process between parties rather than a state obtaining at a moment, and is described by what it produces: shared undertakings, revisable understandings, and the capacity of each party to raise a claim the other must answer.

A *relational condition* is an arrangement whose presence or absence changes which relations can be formed or continued, without determining which of them are. Channels, venues, procedures, standing mandates, and the availability of an intermediary are conditions in this sense.

The *field* of a set of parties is the set of arrangements consistent with those conditions, and *condition governance* is action on the field rather than on the arrangements within it. A *custodian* maintains conditions constituting the field of parties in conflict without directing that maintenance at any particular arrangement within it.

A power is *field-constitutive* when its exercise changes the set of arrangements available to the parties rather than which available arrangement they select. The companion paper's constraint

holds that a custodian may judge, testify, publish and persuade, and may do so asymmetrically, and may not employ field-constitutive powers to secure a substantive outcome. That constraint is carried into this paper unchanged. What this paper adds is the credibility of the undertaking, not a further constraint on its content.

Two terms are used here that the companion paper did not require. *Capture* is the condition in which an actor's field-constitutive decisions come to track one party's preferences. *Non-capture* is the content of the undertaking whose credibility is at issue: an undertaking that the declarant's field-constitutive decisions will not come to track any party's preferences.

The vocabulary belongs to a wider framework of generative relational analysis developed across a connected series of papers. The present paper uses the terms above and leaves the remainder of that framework aside. Its relation to relational ontology in the social sciences, and to the constructionist account of relational being, is addressed in Section 3.8.

3 Literature Review

This section locates the account among the literatures it draws on and states what each already establishes. Several of the paper's working claims are already made, in whole or in part, by work surveyed here; those are conceded where they arise rather than defended. One result surveyed here contradicts a prescription the paper makes, and it is recorded as a contradiction.

3.1 Modes of Trust Production

Zucker's account of how trust was produced in the American economy between 1840 and 1920 supplies the framework this paper depends on [69]. She distinguishes three modes. *Process-based* trust rests on a record of past or expected exchange between the parties. *Characteristic-based* trust rests on shared background, such as common origin or membership. And *institution-based* trust rests on formal structures external to both parties, including professional certification, regulation, and intermediaries. Her historical argument is that immigration, internal migration, and the instability of firms eroded the first two and forced a shift to the third.

The transposition is direct. A custodian dealing with strangers has no shared record and is often chosen precisely because it shares no background with either party, so the first two modes are unavailable to it by construction. That leaves the third, which is the mode an office has and an isolated individual does not. The paper's problem can therefore be stated in Zucker's terms: how an individual obtains institution-based trust without an institution.

Adjacent work supplies the components of an assessment. Mayer, Davis and Schoorman model perceived trustworthiness as ability, benevolence and integrity, hold that the truster's own propensity dominates where information is absent, and find that integrity weighs most heavily early in a relationship while benevolence grows in weight over time [43]. Rousseau, Sitkin, Burt and Camerer supply the definition this paper uses, treating trust as the intention to accept vulnerability on the basis of positive expectations of another's intentions or conduct [54].

Two findings bound what may be inferred. The investment game of Berg, Dickhaut and McCabe establishes that anonymous strangers transfer substantial sums with no contractual or reputational protection [6], and the result is robust across a large replication base with regional variation [32, 49]. What this establishes is that trust between strangers occurs; it does not establish that a stated commitment is believed, which is the present question. Meyerson, Weick and Kramer's account of swift trust in temporary groups is nearer, and shows that strangers who must act together import expectations from roles and categories, proceed as though trust were present, and calibrate

afterwards [44]. That mechanism is available to a custodian who occupies a recognised role, and it is unavailable to one who does not.

A different line of work states the objection this paper must answer, and it is recorded here at length because the remainder of the paper proceeds despite it. Baier holds that trust is reliance on another's goodwill toward one, and marks the feature that distinguishes it from mere reliance: trust can be *betrayed* and not merely disappointed [4]. Jones develops the account as an attitude of optimism about the trusted party's goodwill and competence, such that the trusted party is moved directly and favourably by the thought that they are being trusted [33], and later examines trustworthiness in its own right [34]. Hawley substitutes a commitment for goodwill, treating trust as reliance on another to meet a commitment they have [29].

On Baier's and Jones's accounts, what this paper constructs is not trust. A party who keeps an undertaking because a third party determines breach and because breach would be expensive is not moved by the thought of being trusted; what they supply is reliability, and a counterparty who relies on them is not exposed to betrayal but only to loss. The objection extends further. If credibility can be engineered, then the capacity to engineer it is itself evidence of nothing about goodwill, and a counterparty aware of the machinery has reason to ask why the machinery was built.

Two replies are available and neither disposes of the objection. The paper concerns credibility rather than trustworthiness, and says so in its front matter and in Section 5.1; a counterparty deciding whether to enter a relation with a custodian may reasonably want reliability and not goodwill, since the undertaking concerns the administration of conditions rather than care for either party. And Hawley's commitment account is compatible with the construction here, since a commitment can be relied on without goodwill being ascribed. What survives is the residue that matters most: an account of how commitments are made believable is available to a party who intends to defect and is describing, in advance, the appearance they would need to construct. The paper has no defence against that use and states it in the front matter as a request rather than as an answer.

Gambetta's study of criminal communication is the closest empirical literature to the paper's problem, because it concerns parties who must convey trustworthiness precisely where institutions are unavailable and where everyone has reason to lie [22]. His finding is that the solutions rest on signals that are costly and hard to fake, and his companion study of taxi drivers examines the same assessment made rapidly and with little information [23]. The present paper's condition on declarations, developed in Section 5.3, is an instance of this and is conceded as such.

3.2 Costly Signalling and Credible Commitment

Two results establish the mechanism by which a statement about oneself becomes informative.

Spence's model of job-market signalling shows that a signal separates types only where it is sufficiently more costly for the type that would misrepresent itself; where the cost is equal, every type sends the signal and it conveys nothing [59]. The condition is on the cost structure rather than on the content of the signal, and it is indifferent to sincerity.

Williamson's account of credible commitment locates the same structure in exchange, holding that credible commitments and credible threats appear mainly in conjunction with irreversible and specialized investments, and treating hostages, understood as sunk relationship-specific assets, as the device by which a party makes its own defection expensive [68].

Taken together these establish that a declaration is believed on the strength of what it would cost to make falsely, and that the cost must be borne irreversibly rather than promised. The paper's claim that a declaration must be differentially costly adds nothing to this and is conceded to it. What remains open, and what Section 5 takes up, is which costs are available to an individual who holds no specialized assets that a counterparty values.

3.3 Precommitment and the Limits of Self-Binding

Elster's two treatments of precommitment bracket the position this paper must occupy. The earlier work developed self-binding as a rational response to anticipated weakness, taking Ulysses at the mast as its figure [16]. The later work substantially retracts the extension of that figure to collective self-binding [17]. Elster there endorses Seip's observation that in politics people never try to bind themselves but only to bind others, and notes that Ulysses did not only bind himself: he also put wax in the ears of the rowers, so that the arrangement rested on the control of other parties rather than on self-restraint alone.

This bears directly on the paper's central claim and largely anticipates it. Section 7.2 therefore opens with the concession and confines itself to what the individual, non-capture case adds.

Sunstein and Ullmann-Margalit supply the nearest existing taxonomy of the devices in question [60]. They treat agents as adopting second-order strategies to reduce the burden and the risk of error in first-order decisions, and enumerate rules, presumptions, standards, delegation of authority to others, proceeding by small steps, and picking rather than choosing. They classify these by cost structure, distinguishing strategies that impose high costs before the decision from those that impose low costs throughout, and from those that impose low costs before the decision while exporting the high costs at the time of decision to others.

The overlap with Section 7.3 requires a statement. Delegation appears in both treatments, and the category of strategies that export costs to others is adjacent to what this paper calls ceding authority. The classifying axis differs. Sunstein and Ullmann-Margalit classify by the cost and error of the agent's own future decisions, and the present paper classifies by credibility to an audience that was not party to the decision. A strategy that is optimal on their axis may be worthless on this one, since a rule adopted to relieve oneself of deliberation persuades nobody of anything. The trilemma is therefore not a redescription of their taxonomy, and it is offered as an addition to it rather than as a replacement.

3.4 Self-Legislation and the Bootstrapping Objection

The philosophical literature on self-legislation states the paper's central difficulty in its sharpest form, and does so without reference to credibility at all.

Korsgaard grounds obligation in practical identity and in the constitutive standards of agency, and names the difficulty that attends any such grounding as the paradox of self-constitution: action both expresses and constitutes the self, so that the agent must already be present in order to choose while also being made by the choosing [35, 36]. The vocabulary of self-constitution is hers, and Section 7.1 uses a related term for a different object; the difference is stated there.

O'Shea presses the objection this paper must answer. Examining Kantian constructivism in Korsgaard and O'Neill, he argues that neither succeeds in defending self-legislation as the fundamental source of normativity without that legislation collapsing into arbitrariness [50]. The structure of the objection is that an agent who authors a law may repeal it, so that nothing is bound by it. Seeman reaches a related conclusion, holding that the paradox generates an arbitrariness that undermines the formation of moral laws [57].

The objection is the philosophical form of the observation from which this paper proceeds, that a rule whose breach the rule-maker determines will not be believed. The paper does not attempt to resolve the moral-philosophical question of whether self-legislation can ground obligation. It takes the narrower question of whether a self-stated rule can be made credible to a third party, and answers it by relocating the determination of breach, which is a route the moral-philosophical literature has reason not to take, since relocating the determination of obligation to others is precisely what

autonomy-based accounts exist to avoid.

3.5 Diachronic Agency and the Binding of a Later Self

A self-stated rule is stated at one time and kept at others, and a literature on the agency of planning creatures addresses what holds it across the interval.

Bratman's account treats intentions as states that are neither reducible to desires and beliefs nor readily abandoned, and that exist because agents with limited cognitive resources must coordinate with their own futures [8, 9]. Holton develops the case that matters here. He distinguishes, among intentions, the *resolution*: an intention formed in anticipation of contrary inclination, whose function is to hold firm against it [30]. A resolution is on his account an intention not to reconsider the underlying intention when the anticipated temptation arrives, and he argues that declining to reconsider can be rational even where the agent acquires no new information in the interval, which is a stronger claim than Bratman's.

The relevance is exact and its limit is equally exact. An undertaking of non-capture is a resolution in Holton's sense, since it is formed in anticipation of pressure and its whole function is to hold when the pressure arrives, and his account supplies the reason why keeping it may be rational even at a moment when abandoning it appears best. What the account addresses is the agent's own rationality in keeping the resolution. It does not address whether a party who was not present has reason to expect the agent to keep it, which is the question this paper takes. The two are related in an unhelpful direction: an account explaining why a resolute agent is rational supplies no means by which an observer distinguishes a resolute agent from an irresolute one.

Parfit's treatment of personal identity presses a further difficulty [52]. If what matters in the persistence of a person is psychological continuity and connectedness, and if these admit of degree, then the force with which a later self is bound by an earlier self's undertaking is itself a matter of degree, and diminishes as the interval lengthens.

This supplies an argument for the paper's central move that Section 7 reaches from another direction. If the binding force of a self-undertaking depends on the continuity between the undertaking self and the acting self, then self-binding is weakest precisely where the undertaking is oldest, which is where a counterparty most needs it to hold. A rule whose determination has been relocated to others is indifferent to that continuity, since the parties determining breach are not the party whose identity over time is in question. The relocation therefore addresses a difficulty in the metaphysics of the committed party as well as a difficulty in the epistemics of the observing one.

3.6 Enforcement in the Absence of a Sovereign

A body of historical and empirical work establishes that individuals sustain enforceable commitments where no state enforces them, and it does so in enough detail to identify the mechanism.

Milgrom, North and Weingast examine the medieval Law Merchant and show that private judges holding no power to compel nonetheless supported honest trade, by maintaining and transmitting information about whether a merchant had satisfied judgments against him [46]. Enforcement was decentralized: a merchant who ignored a judgment could be identified by any counterparty who consulted the record, and was thereafter avoided. The institution supplied not a sanction but the information on which many private sanctions could be based.

Greif's studies of the Maghribi traders identify a related structure, in which a multilateral punishment strategy makes honesty self-enforcing because an agent who cheats any member of the coalition is thereafter hired by none [27, 28]. The historical accuracy of that account is contested;

Edwards and Ogilvie reappraise the evidence and dispute the reading [14]. The dispute concerns whether the Maghribi case exhibits the mechanism, and leaves the mechanism itself untouched.

Bernstein documents a modern instance in which participants deliberately opt out of state enforcement, showing that the diamond industry resolves disputes through private arbitration, reputation and community sanction rather than through the courts available to it [7]. Ellickson documents the same in a setting with no trade association at all, finding that neighbours settle disputes by informal norms while remaining substantially ignorant of the law that governs them [15]. Ostrom's analysis of self-governing communities identifies monitoring and graduated sanctions among the conditions under which such arrangements endure [51].

This literature bears on the paper in two opposite directions and both are recorded. It supports the claim that enforcement of a stated rule is communal rather than self-administered, since in none of these cases does the committed party determine its own breach. And it presses against the premise from which the paper begins, since these are individuals sustaining credible commitments without a sovereign. Section 5.2 accordingly states the three deficits as deficits of the isolated individual rather than of the individual as such, which is the correction this literature requires.

3.7 Role Morality and Professional Independence

Applbaum's treatment of professional roles supplies a caution the paper must observe. He examines the claim that occupying a role makes permissible what would otherwise be wrong, and argues that institutions ordinarily cannot mint moral permissions [2]. The caution applies wherever this paper grants a custodian latitude that a private person would lack.

The response is that the paper's claims concern credibility rather than permission. What makes an undertaking believable is a separate question from what the undertaking may license, and the constraint on a custodian's instruments is inherited from the companion paper rather than generated by the role. Where the two questions meet, in Section 8, the caution is applied rather than set aside.

The regimes described in Section 2.1 constitute the practical literature for this section. Article 100 of the Charter and the associated Staff Regulations are notable for combining three devices that the present account treats separately: an individual declaration, an external witness to it, and a reciprocal obligation on the parties who might otherwise seek to capture the declarant [62, 63]. No instrument available to a private individual combines all three, and Section 8.4 takes that as its principal finding.

3.8 Relational Ontology and the Generative Relational Framework

The framework this paper works within treats relations as prior to the parties they relate, and that commitment has an established literature which must be distinguished from it.

Emirbayer's programmatic statement for a relational sociology holds that the social world consists in dynamic, unfolding relations rather than in static substances, and distinguishes a transactional view, in which the terms of a relation derive their meaning from the relation, from an interactional view in which pre-formed units act upon one another [18]. The companion study of agency develops the same commitment for action [19]. The present framework shares the priority claim and takes it as background rather than as a contribution.

Two terminological collisions require explicit statement. Gergen's account of relational being treats the self as constituted in and through relationship rather than as a bounded unit that subsequently enters relations [24]. The name of the present author's framework stands one word from the title of that work, and the objects differ: Gergen theorises selfhood and its social constitution, whereas the framework used here concerns the governance of the conditions under which relations

Table 1: Antecedent literatures and contribution boundaries

Literature	Licensed role	P002 boundary
Trust production	Three modes of trust production and the shift to institutional sources [69]; components of assessment [43]; trust between strangers [6]; swift trust from roles [44]	Supplies the problem statement. How an individual obtains institution-based trust without an institution is left open by it.
Costly signalling	Separation requires differential cost [59]; credible commitment requires irreversible specific investment [68]; signalling where institutions are absent [22]	The condition on declarations is conceded entirely. Which costs are available to a person holding no valued specific assets is not addressed there.
Precommitment	Self-binding and its collective limits [16, 17]; the taxonomy of second-order strategies [60]	Classified by decision cost rather than by credibility to an audience; the trilemma is offered on the second axis.
Self-legislation	The paradox of self-constitution and the arbitrariness objection [36, 50, 57]	States the difficulty and declines the remedy this paper adopts, for reasons internal to autonomy-based accounts.
Diachronic agency	Resolutions as intentions formed against anticipated temptation [30, 8]; degrees of psychological continuity [52]	Addresses the agent's own rationality in keeping a resolution, and supplies no means by which an observer distinguishes the resolute from the irresolute.
Private ordering	Communal enforcement without a sovereign [46, 28, 7, 15, 51]	Establishes the mechanism and corrects the paper's premise; the individual outside such a community is not its subject.
Role morality	Institutions ordinarily cannot mint moral permissions [2]	Concerns permission; the present claims concern credibility.
Relational ontology	Priority of relations over the parties related [18, 24]	Taken as background; the object here is the credibility of an undertaking.

continue to be generated. The resemblance of terms is stated so that it is not mistaken for derivation. Separately, the term generativity carries an established meaning in developmental psychology, where Erikson uses it for a stage of adult concern with establishing and guiding the next generation, in a body of work that also supplies the notion of basic trust on which later trust research draws [20]. Neither usage is the one intended here, and no continuity with either is claimed.

3.9 Boundary of the Present Contribution

Table 1 records what each surveyed literature licenses and where the present contribution begins.

Four positions are left unoccupied by the literatures surveyed. No treatment located here distinguishes the registers through which trust is produced in a way that assigns each a distinct requirement and a distinct failure condition. No treatment states the joint unavailability of credibility, revisability, and retained authority as a constraint on commitment devices, though a trilemma of the same form is established for optimal auctions over entirely different properties [1] and is conceded in Section 7.3. None derives, from the length of an observation record, a difference in the volatility of individual and institutional reputations, or the design consequence that follows from it. And none treats the community recruited to enforce a commitment as itself a surface on which capture occurs. The claim is that these four are unoccupied, not that their components are unprecedented;

the components are conceded above. A systematic originality audit remains outstanding and is recorded in Section 11.

4 Method and Case Selection

This section states what the comparison is asked to do, why these four instruments were chosen, and what would count against the account.

4.1 Comparative Instrument Study

The paper develops a conceptual account and examines four instruments by which individual independence is or might be manufactured. The instruments are compared against a common set of questions: what the instrument requires the committed party to give up, who determines whether the commitment has been breached, what a stranger can check without the committed party's cooperation, and what the instrument leaves revisable.

A comparison of this kind can establish that a distinction has purchase, can show that instruments differing in domain resolve the same trade-off in different ways, and can expose an instrument that resolves it badly. It cannot establish that the instruments work, since none of the four is evaluated here against evidence of its effects, and it cannot establish that the set is exhaustive.

Established description is separated from analysis, as in the companion paper. Section 2.1 records what the regimes provide and Section 8 records what the account makes of them.

4.2 Selection of the Four Instruments

The four were chosen to vary along the dimension the account identifies rather than to represent a domain.

The international civil service places the undertaking inside an office and supplies external enforcement, and it is included because it is the developed case against which any private instrument must be measured. Judicial recusal and the blind trust remove a capacity rather than binding a will, and are included because they occupy a different corner of the trade-off developed in Section 7.3. Auditor independence relies substantially on disclosure and appearance, and is included because Section 7.4 argues that disclosure carries a specific hazard. And the author's own instrument is included because the paper would otherwise theorise a class of device without examining a member of it that can be described from the inside.

The fourth selection carries an interest, disclosed in the front matter. Its inclusion is defensible only if it is evaluated by the criteria applied to the other three and reported as failing where it fails, and Section 8.3 is written accordingly.

Two absences are recorded. Certification and licensing regimes, by which individuals borrow credibility from bodies that hold reputations of their own, belong to the same class and are omitted for reasons of space rather than of principle. And the private-order communities examined in Section 3.6 are not treated as instruments here, because membership of such a community is a condition rather than a device.

4.3 Conditions of Disconfirmation

The account should be narrowed or withdrawn under any of the following conditions.

First, if the three deficits identified in Section 5.2 are shown to be consequences of isolation with no residue attaching to the individual position as such, then the paper's problem dissolves into the general problem of belonging to a community.

Second, if an instrument is exhibited that holds credibility, revisability and retained authority together, the trilemma of Section 7.3 is false.

Third, if the registers distinguished in Section 6 collapse into one another under examination, so that securing one reliably supplies the others, the distinction does no work and should be removed.

Fourth, if reputational volatility is shown to be independent of the length of the observation record, the fragility asymmetry of Section 9.1 is false.

Fifth, if the devices this paper recommends are shown to worsen the conduct they govern, the account is not merely incomplete but harmful. Section 7.4 treats evidence of exactly this kind and does not dispose of it.

5 The Credibility Problem for an Individual Custodian

This section states the problem the remainder of the paper addresses. It specifies what must be made credible, identifies what the individual position lacks, states the condition a declaration must satisfy, and notes that the value of satisfying it depends on which market the declarant is addressing.

5.1 The Commitment an Individual Custodian Must Render Credible

The undertaking at issue is narrow and it is worth stating precisely, because the difficulty of making it credible follows from its content.

A custodian does not undertake to be impartial in judgment, to treat the parties identically, or to refrain from forming and expressing views. It undertakes that its field-constitutive decisions will not come to track any party's preferences: that access, hosting, procedure, channel and agenda will be administered without regard to which party benefits. The undertaking is one of non-capture in the sense defined in Section 2.3.

Three features of this undertaking make it hard to evidence. It concerns the grounds of decisions rather than their content, and grounds are not observable. It concerns a disposition over an indefinite future rather than a discrete act. And its satisfaction is largely invisible, since a custodian that has not been captured looks exactly like a custodian that has not yet been approached. What a stranger can observe is a record of decisions, from which the grounds must be inferred; and the inference is weak precisely where the custodian has not yet faced a serious attempt at capture.

5.2 Three Deficits of the Isolated Individual Position

The comparison of Section 2.2 can now be stated as the paper's diagnosis. Relative to an office, an individual undertaking lacks duration, observability, and cost of exit. Each corresponds to something a counterparty relies on when it credits an institutional undertaking, and each is absent for reasons independent of the individual's sincerity.

The literature surveyed in Section 3.6 forces a qualification on this diagnosis, and the qualification is adopted rather than resisted. Diamond merchants, ranchers, medieval traders and the participants in Gambetta's study sustain credible commitments without any sovereign [7, 15, 28, 22]. They do so because a community supplies what the individual lacks: it persists beyond any member, it observes and records conduct, and it can exclude, which makes departure expensive. The deficits are therefore deficits of the *isolated* individual rather than of the individual as such.

That correction does not dissolve the problem and it changes its shape. If the three properties are supplied by a community rather than by the individual, then the question is no longer how an individual generates credibility but how an individual comes to stand within a structure that supplies it, and on what terms. Two consequences follow and are developed later. The remedy is necessarily relational, which Section 7.2 states as a general claim about commitment devices. And the structure that supplies the properties acquires authority over the individual who depends on it, which Section 9.3 treats as an exposure rather than as a solution.

A residue attaches to the individual position even so. A community supplies these properties to its members in respect of conduct the community can observe and cares about. A custodian's undertaking of non-capture concerns conduct toward parties who are typically outside any community the custodian belongs to, and it is not obvious that a professional or local community either observes or penalises capture by an external party. The deficits are thus mitigated by membership and not necessarily removed by it.

5.3 Differential Costliness and the Condition of Credibility

The condition a declaration must satisfy is supplied by the literature and is conceded to it. Spence's separation result holds that a signal conveys information only where it is sufficiently more costly for the type that would misrepresent itself, so that a signal available at equal cost to all types is sent by all and distinguishes none [59]. Williamson locates the same structure in the commitments parties make to one another, holding that credibility attaches to irreversible and specialized investment rather than to statement [68].

Applied here, a declaration of non-capture is informative only if making and keeping it is more expensive for a declarant who intends to remain capturable than for one who does not. A declaration that costs nothing is sent by both types and separates neither.

The question this leaves is which costs are available. A state posts territory, institutions and accumulated standing. A merchant posts membership of a trade whose members can exclude him. An individual custodian who holds no office, no specialized assets that a counterparty values, and no membership that can be withdrawn has, at first inspection, nothing to post.

Three candidate costs are available and each has a limitation. *Foreclosure* costs are incurred by undertaking not to accept positions that would be available otherwise, which is expensive in proportion to how attractive those positions were and is verifiable only by observing that they were declined. *Sunk relational* costs are incurred by building arrangements that others come to rely on, which is the individual analogue of the accumulation described in the companion paper, and which takes years. *Recruited* costs are incurred by placing the determination of one's own breach in other hands, which is immediate and which transfers authority. Section 7 argues that the third is not one option among three but the structure underlying the other two.

5.4 Demand for Neutrality in Principal-Neutral and Principal-Agent Markets

A property that makes a person valuable in one market may reduce their value in another, and the undertaking at issue appears to have this character.

Where a role requires acceptability to parties who do not trust one another, the absence of alignment is the qualification. Mediation, arbitration, standard-setting, protecting-power work and the international civil service all select for it, and Article 100 makes the requirement explicit by prohibiting the acceptance of instruction from any government [62]. Call these principal-neutral markets.

Where a role requires an agent to advance one party's interest against others, the same absence reads as a deficiency. Advocacy, litigation, national civil service and security work select for reliable alignment, and a candidate who has undertaken never to align is, on the face of it, a candidate who has undertaken not to perform the role. Call these principal-agent markets.

The conjecture is that an undertaking of non-capture is a positive signal in the first and a negative signal in the second, and that the same properties produce both effects.

The criteria applied in one principal-agent market support the second half of the conjecture, and they do so in terms close to the paper's own. Eligibility for access to classified information in the United States is assessed under the National Security Adjudicative Guidelines, whose factors include allegiance, whether the individual has conflicting allegiances or divided loyalties, and vulnerability to pressure or manipulation by a foreign government [53]. Departmental guidance states the requirement affirmatively: an individual must demonstrate unquestioned allegiance to the United States and preference for it over any other country, and where this cannot be established a clearance is refused [65]. A demonstrated preference for one state is thus the qualifying condition in that market, and an undertaking never to hold such a preference is its negation.

Two qualifications belong with this. The guidelines assess conduct rather than status, and the same sources record that dual nationality is not in itself disqualifying and that adjudication proceeds on a whole-person basis [53]. And the material establishes what the criteria require, not what happens to candidates. No study of the employment consequences of declared non-alignment was located in the survey underlying this paper. The conjecture is therefore supported at the level of stated criteria on the negative side, by Article 100 on the positive side [62], and by no evidence at all at the level of outcomes. Section 11 records it accordingly.

6 Registers of Trust Production

Trust is treated in most of the literature surveyed in Section 3.1 as a single quantity with several sources. This section proposes instead that trust is produced through distinct registers, that each carries its own requirement and its own characteristic failure, and that an instrument securing one supplies little of the others. The distinction is put to work in Section 8, where the four instruments are found to operate in different registers.

The three registers are named after Lacan's division of the psychical field into symbolic, real, and imaginary orders, and the borrowing requires a statement of its limits. What is taken is the division and the account of what each order can and cannot do. What is not taken is the clinical apparatus, the theory of the subject, or any claim that trust between institutions is continuous with transference. The registers are used here as a typology of mechanisms, and the typology stands or falls on whether the three mechanisms have distinct requirements and distinct failures, which is a question that can be settled without adjudicating the psychoanalytic theory.

6.1 Trust Produced through the Symbolic Order

In the first register, an undertaking is credited because an order of signs stands behind it. A title, an office, a form of words, a signature, a credential: what is trusted is not the person but the position the person occupies within a system of recognized places, and the system rather than the person supplies the guarantee.

Lacan's account of speech supplies the mechanism in its purest form. He distinguishes empty speech from full speech, the latter being speech in which the speaker is engaged by what is said, so that the utterance binds in the manner of a vow or a contract rather than merely reporting [39].

The founding of a symbolic pact is an act of speech of this kind, and its force derives from the order within which it is uttered rather than from the sincerity of the utterer.

Two features of that order matter here. Its guarantee is positional rather than evidential: the holder of an office is credited without the crediting party inspecting the holder's conduct, which is what makes the register efficient. And the order that guarantees is not itself guaranteed. Fink's exposition records the point in Lacan's terms as the barred status of the Other: the symbolic order is the position from which meaning is guaranteed, and there is no further position guaranteeing it [21]. Schrans traces the movement in Lacan's own development by which the guarantee shifts from imaginary identification to the symbolic order [56].

The characteristic failure of the register follows from the second feature. If the guarantee rests on recognition rather than on evidence, then it fails when recognition weakens, and it fails all at once rather than by degrees, since nothing beneath it was carrying weight. Žižek's account of ideology turns on the lack in the guaranteeing order and on the forms of belief that persist around it [66], and his discussion of the decline of symbolic efficiency describes the condition in which the order's guarantees are no longer effective even where they are still uttered [67].

The register has been examined in the same terms outside psychoanalysis proper. Work on the psychoanalytic reading of law treats legal authority as resting on a projected singularity and permanence that masks the plurality beneath it [25], and related work examines how legal orders legitimate themselves and what their guarantees conceal [55]. These are cited as evidence that the register's structure has been analysed for institutions rather than as endorsements of those analyses.

An individual custodian without an office holds nothing in this register. This is the precise sense in which Zucker's institution-based trust [69] and Luhmann's system trust [41] are unavailable to the isolated individual, and it is the deficit that Section 8.1 finds the international civil service supplying and Section 8.3 finds a private instrument unable to supply.

6.2 Trust Produced through Shared Ordeal and the Unsymbolized

In the second register, an undertaking is credited because of what has already been seen. Parties who have been through something together know things about one another that no account would have conveyed, and the knowledge is held in the form of having been present rather than in the form of a statement.

The evidential structure of this register is worth stating precisely, because its strength has a specific source. What is informative is conduct under conditions where defection was available and attractive. Compliance where nothing was at stake conveys almost nothing, which is why duration alone is a poor proxy: a long relation with no occasion for betrayal supplies few observations of the relevant kind. This is Spence's separation condition [59] restated in the language of experience rather than of signalling, and it is why Gambetta finds costly and hard-to-fake signals carrying the weight where institutions are absent [22].

The register is the one in which process-based trust in Zucker's sense is produced [69], and it is the strongest of the three for the parties who hold it. It has two characteristic limitations. It is *slow*, because occasions of the relevant kind cannot be scheduled without manufacturing them, and manufacturing them is a distinct wrong that Section 11 records. And it is *non-transferable*: what one party knows of another through having been present cannot be conveyed to a third party who was not, and any attempt to convey it converts it into a statement in the first register, where it is worth what the speaker's position is worth rather than what the experience established.

The non-transferability is the structural fact this paper most needs. It explains why a custodian who has been thoroughly tested by two parties is still unknown to a third, and why the problem of individual credibility does not diminish with experience in the way that competence does.

6.3 Trust Produced through Identification and Image

In the third register, an undertaking is credited because the crediting party recognizes itself in the one who gives it. Shared origin, shared formation, shared manner, a recognizable resemblance: the party is trusted because it is read as being of the same kind.

This is characteristic-based trust in Zucker's sense [69], and it is the mode her account describes as eroding under migration and mobility. Its efficiency is considerable, since it requires neither a record nor an institution, and Meyerson, Weick and Kramer's account of swift trust turns substantially on the categorical expectations it makes available [44].

Its characteristic failure is that it tracks resemblance rather than conduct. A party that resembles the truster is credited whether or not it is trustworthy, and a party that does not is discredited on the same basis, so the register produces both misplaced confidence and unwarranted exclusion from the same mechanism. Applied to custodianship the difficulty is acute in a further way: a custodian is often chosen precisely because it belongs to neither party's kind, so a custodian that succeeds in this register with one party has by that fact given the other a reason for distrust. The register is therefore not merely weak for a custodian but structurally adverse, which distinguishes it from the first two.

6.4 Trust Produced across Registers

The three registers are not degrees of one quantity, and the differences can be stated as a set of contrasts. The first is fast, transferable, positional, and fails all at once. The second is slow, non-transferable, evidential, and fails only if the evidence is misread. The third is fast, partially transferable through shared category membership, resemblance-based, and adverse for a custodian by construction.

Three consequences follow for the design of an instrument.

Registers substitute poorly. A custodian who holds standing in the first and nothing in the second is believed until the guaranteeing order weakens, at which point nothing remains; a custodian who holds the second and nothing in the first is trusted deeply by a few and unknown to everyone else. Neither condition is a partial version of the other.

An instrument may convert between registers at a cost. Witnessing converts what is held in the second register into a statement in the first: a party who was present attests, and the attestation is thereafter worth what the attesting party's position is worth. The conversion is what makes the second register portable and it is also lossy, and the loss is systematic rather than random, since what survives conversion is what can be stated in the terms the first register recognizes.

And the registers fail independently, which is a resource rather than only a complication. An instrument that stands in more than one register does not fail when one weakens. Section 8.4 finds this to be the principal advantage of the international civil service over the alternatives, since Article 100 and the associated declaration combine a positional guarantee with a witnessed act and a reciprocal obligation on the parties who might capture the declarant [62, 63].

6.5 The Registers and the Generative Relational Account

Two points connect this typology to the framework within which the paper works.

The registers concern what a relation can carry rather than what a person is. Trust in the second register is not a property of the trusted party but a product of a history the parties made together, and its non-transferability follows from that: it is held by the relation rather than by either party, so neither can take it elsewhere. This is the priority of relations over relata that relational sociology

asserts [18], applied to a particular object; it is not an additional claim about selfhood of the kind Gergen develops [24].

A third connection is to a debate in which neutrality has been examined from within a relational commitment, and it supplies both support and a caution. Psychoanalysis maintained an ideal of the neutral analyst and then subjected it to sustained criticism as relational accounts of the clinical situation developed [47]. Greenberg argues that neutrality entered the clinical setting from a research setting where it belonged and fails in the clinical one [26]; Nersessian defends a reconstructed version against that line of argument [48]. The debate is not settled and is cited here for what both sides accept rather than for either's conclusion.

What both accept is the position this paper's companion took: that a party standing between others is inside the relation rather than outside it, and that its conduct participates in what it observes. Benjamin's account of thirdness develops the consequence, holding that a relation may be mutual and asymmetric at once, so that recognition runs both ways while role and responsibility do not [5]; Aron examines the same tension between mutuality and asymmetry [3]. That combination is exactly the custodian's position as the companion paper described it, arrived at in a discipline with no connection to the study of conflict. The caution is that the discipline which examined this longest has not resolved whether neutrality is coherent as an ideal, and this paper's confinement of the question to credibility rather than to virtue does not dispose of that difficulty.

And the registers give the companion paper's conditions a second dimension. That paper treated relational conditions as arrangements whose presence changes what relations can be formed. The present typology adds that conditions differ in the register through which they operate: an accredited venue operates in the first, a history of joint work in the second, a shared professional formation in the third. A field composed of conditions in a single register is correspondingly exposed to that register's characteristic failure, which is a consideration in the design of a field that the companion paper did not identify.

7 Self-Legislation and Relational Enforcement

This section develops the paper's central result. It distinguishes three things that are commonly called self-legislation, states what the devices by which individuals bind themselves have in common, derives the constraint that follows, and examines a prescription that the evidence contradicts.

7.1 Self-Discipline, Self-Legislation, and Self-Constitutionalization

Three arrangements are distinguished, and they differ in who determines whether a breach has occurred.

Under *self-discipline*, a person adopts a rule, observes it, and judges their own compliance. The arrangement is entirely internal. It may be efficacious, in the sense that the person does what they resolved, and it conveys nothing to a third party, because the observation on which any judgment of compliance rests is available only to the person judged.

Under *self-legislation* as the term is used here, a person states a rule publicly, so that its content is fixed and available to others, while the determination of compliance remains with the person. The statement adds something to self-discipline, since it makes the rule checkable in principle where conduct is observable. It does not resolve the difficulty, because a person who both states the rule and rules on their own conduct under it has supplied no reason for a stranger to prefer their judgment to any other.

Under *self-constitutionalization*, a person states a rule and places the determination of breach elsewhere, so that some party other than the declarant decides whether the rule has been kept. The

difference is not one of stringency but of who holds the judgment.

Two terminological notes are required. Korsgaard's self-constitution names a different object: the constitution of an agent through action, and the constitutive standards that attend agency as such [36]. The present term concerns an institutional arrangement and carries no claim about the metaphysics of agency. And self-legislation in the Kantian tradition concerns the ground of obligation rather than its enforcement, which is why that tradition faces the arbitrariness objection recorded in Section 3.4 [50]. This paper faces the same objection in its own register: a rule whose breach the rule-maker determines is a rule whose keeping the rule-maker may redefine.

7.2 Recruitment of Others into Enforcement

The general point of this subsection is already established and the concession is made first.

Elster's mature treatment holds, following Seip, that in politics people never try to bind themselves but only to bind others, and observes that Ulysses did not only bind himself but also stopped the rowers' ears, so that even the paradigm case of self-binding rested on the control of other parties [17]. The private-order literature establishes the same for enforcement: in the Law Merchant, the register of unsatisfied judgments made decentralized sanction possible [46]; in the trading coalitions, exclusion by all members enforced honesty on each [28]; in the diamond trade and in Shasta County, community sanction substituted for the courts [7, 15]; and in Ostrom's cases, monitoring by members and graduated sanctions were among the conditions of endurance [51]. In none of these does the committed party determine its own breach.

What the present case adds is the scope of the claim and one consequence. The scope: the literature above concerns communities enforcing rules on members whose conduct the community observes and cares about, whereas the undertaking at issue here concerns conduct toward parties outside any such community, as Section 5.2 recorded. The consequence is a restatement of what self-legislation is.

Claim 7.1 (Relational enforcement)**.** A commitment device renders a self-stated rule credible only by transferring the determination of breach, in whole or in part, to parties other than the declarant. The self supplies the rule and a field supplies the determination. There is accordingly no purely self-imposed credibility, and the term self-legislation names the origin of a rule rather than the arrangement that makes it hold.

Claim 7.1 is consistent with capability-removing devices as well as with delegated ones, and the consistency is worth stating because it appears otherwise. A person who removes their own capacity to defect has not retained the determination of breach; they have made the question unnecessary, and they have done so by an act that others can inspect. The blind trust is credible because a third party holds the assets and the arrangement is a matter of record, not because the officeholder has resolved to be impartial [12]. Removal is therefore a form of transfer, in which what is transferred is the occasion for judgment rather than the judgment itself.

7.3 The Commitment Trilemma of Credibility, Revisability, and Retained Authority

Claim 7.1 restricts the space of available devices, and the restriction has a structure.

Three properties may be wanted of a commitment device. *Credibility* is the property that a party who was not present has reason to believe the rule will be kept. *Revisability* is the property that the rule can be amended as circumstances change, which the framework within which this paper works treats as a requirement rather than a convenience. *Retained authority* is the property that the declarant continues to determine what their own conduct requires.

Claim 7.2 (Commitment trilemma)**.** No commitment device holds credibility, revisability, and retained authority together. Removing the capacity to defect secures credibility and retains authority, and forecloses revision. Declaring a rule while judging one's own compliance secures revisability and retains authority, and lacks credibility. Placing the determination of breach with others secures credibility and revisability, and cedes authority.

The argument is short and follows from Claim 7.1. Credibility requires that the determination of breach not rest with the declarant, by Claim 7.1. That determination may be made unnecessary, by removing the capacity, or relocated, by delegation. Removal operates by foreclosing the conduct and therefore by foreclosing its revision, since a capacity that can be restored at will has not been removed. Relocation preserves the capacity and the possibility of revision under whatever procedure the recipients apply, and transfers to them the authority to say what the rule requires. Retaining both the capacity and the authority leaves the declarant judging their own breach, which by Claim 7.1 is the uncredible case.

A prior result of the same form, and on an adjacent subject, must be conceded before the relation to other work is stated. Akbarpour and Li prove a trilemma for optimal single-item auctions in which only the winner pays: no such auction is simultaneously static, strategy-proof for the bidders, and credible, where a mechanism is credible when the party running it has no incentive to deviate from the rules it has announced; and taking the properties two at a time uniquely characterizes the standard auction formats [1]. The overlap is more than terminological. Their notion of credibility is the present paper's problem in a formal setting, since the question is whether the party who administers a rule can be relied on to administer it rather than to depart from it undetected.

Three differences separate the results. The properties differ entirely, since theirs are properties of a mechanism's form and information structure and the present ones are properties of the arrangement by which a commitment is enforced. Their result is a theorem within a specified class of mechanisms, whereas Claim 7.2 is an argument from Claim 7.1 and is correspondingly weaker. And their credible mechanisms achieve credibility by removing the administrator's opportunity to deviate undetected, which in the present vocabulary is the capability-removing corner, so their result concerns the trade-offs that remain once that corner is occupied. The present claim is offered as an analogue in a different domain and not as an extension of theirs.

The relation to Sunstein and Ullmann-Margalit's taxonomy of second-order strategies [60] was stated in Section 3.3 and is repeated in short form here. Their categories are classified by the cost and error of the agent's own future decisions; Claim 7.2 classifies by credibility to a party who was not present. Delegation appears in both, and their category of strategies that export costs to others is adjacent to ceding authority, so the overlap is real and partial. What does not appear there is the impossibility itself.

Three consequences are recorded. A framework that requires revisability, as this one does, is barred from the capability-removing corner and is therefore committed to ceding authority; the cession is an entailment rather than a regrettable cost. An instrument may be located by which corner it occupies, and Section 8 so locates the four. And a device that appears to hold all three is either not credible, not revisable, or has ceded authority without saying so, which supplies a diagnostic that Section 8.3 applies to the author's own instrument.

7.4 Verifiability, Artifacts, and the Backfire of Disclosure

Claim 7.2 says where the determination of breach must sit. It says nothing about what the parties holding it can observe, and a determination that cannot be made cheaply will not be made.

Two design requirements follow directly. A rule must be stated in terms whose breach is observable, so that a rule against accepting instruction from a government is checkable in a way that a rule

to maintain relational integrity is not. And the conduct governed should generate durable traces, so that the check requires the declarant's cooperation as little as possible. The Law Merchant supplies the model: what made decentralized sanction possible was a maintained record of who had satisfied judgments [46].

The evidence bearing on the second requirement is adverse and this subsection is written around it.

Cain, Loewenstein and Moore examined the disclosure of conflicts of interest and found effects running against the intention of the practice [10, 11]. Recipients of advice discounted biased advice insufficiently even when the bias was disclosed. And disclosure altered the conduct of the advisers, who exaggerated further when their interest had been declared, an effect the authors attribute to moral licensing and to strategic anticipation of the recipient's discounting. The disclosure therefore increased the bias while failing to correct for it.

The blind trust exhibits the same structure in an institutional form. The device is presented as removing the officeholder's knowledge of their holdings, and the standing criticism is that the officeholder knows what was placed into the trust, so that the arrangement functions as an appearance rather than a constraint [12]. What the artifact then supplies is a publicly citable fact of compliance that does not correspond to a constraint on conduct.

Two conclusions follow and the first is a retraction. The prescription that commitment rules should generate artifacts is not sound as stated, because an artifact that certifies compliance without constraining conduct may improve the declarant's standing while worsening what the standing is taken to guarantee. The prescription survives only in a narrower form.

Claim 7.3 (Condition on artifacts)**.** An artifact contributes to credibility only where it is costly to produce falsely and verifiable by a party other than the declarant. An artifact that is cheap to produce and that certifies compliance without exposing conduct to independent check transfers credibility to the declarant while transferring no constraint, and may worsen the conduct it reports.

Claim 7.3 is a restriction on Claim 7.1 rather than an exception to it. The parties recruited must be able to determine breach; an artifact that reports compliance without permitting that determination has not recruited them but has instead supplied them with a substitute for judgment. The distinction is between an artifact that carries evidence and an artifact that carries a claim, and in the register vocabulary of Section 6 it is the distinction between a trace that survives in the second register and an utterance that operates in the first.

8 Case Study: Instruments of Individual Independence

This section examines four instruments against the questions stated in Section 4.1: what the instrument requires the committed party to give up, who determines breach, what a stranger can check without the committed party's cooperation, and what remains revisable.

8.1 The International Civil Service

The regime described in Section 2.1 combines three devices that this paper has treated separately, and the combination is its distinguishing feature.

The rule is stated. Article 100 provides that the Secretary-General and the staff shall not seek or receive instructions from any government or from any other authority external to the Organization, and shall refrain from action reflecting on their position as officials responsible only to the Organization [62]. The prohibition is specific enough that its breach is in principle observable, which satisfies the first requirement of Section 7.4.

The rule is undertaken individually and witnessed. Each staff member makes a written declaration, witnessed by the Secretary-General or an authorized representative, promising to regulate their conduct with the interests of the Organization only in view and neither to seek nor to accept instructions from any government or other external source [63]. In the vocabulary of Section 6.4, the witnessing converts a private undertaking into an act within the symbolic register, where it carries the weight of the position of the party who witnessed it.

The parties who might capture the declarant are themselves bound. Article 100's second paragraph commits each member state to respect the exclusively international character of the staff's responsibilities and to refrain from seeking to influence them [62]. This has no counterpart in the other three instruments and it addresses the problem from the side the others leave open, since the other three constrain only the party who might be captured.

Located against Claim 7.2, the regime occupies the delegated corner. Breach is determined by the Organization rather than by the staff member; the arrangement is revisable, since regulations and rules are amended; and authority over what the undertaking requires rests with the employer. The cost is exactly the one the trilemma predicts. An international civil servant who disputes the Organization's construction of their duty has no independent standing from which to do so, and the independence secured against states is purchased by dependence on the institution.

8.2 Recusal, Blind Trusts, and Auditor Independence

Three professional devices resolve the trade-off differently from one another, which is why they are grouped rather than treated as one.

Recusal operates on the occasion rather than on the disposition. The officeholder is removed from the particular matter in which an interest arises, and nothing is asked of their conduct in matters from which they are not removed. The device is an instance of the maxim that no one should judge in their own cause, and its economy is considerable: it requires no undertaking about the future and no assessment of anyone's state of mind. Its limitation is that it presupposes the interest can be identified in advance and that the matters are separable, neither of which holds for a custodian whose undertaking concerns a standing disposition across an indefinite set of future occasions.

The blind trust operates by removing a capacity, and it is the clearest instance of the capability-removing corner of Claim 7.2. Assets are placed with a trustee so that the officeholder cannot know what they hold and therefore cannot be influenced by holdings they cannot identify. The device is credible in the way that removal is credible, and the standing criticism is that the removal is incomplete: the officeholder knows what was placed into the trust, so that the arrangement is described by critics as an appearance rather than a constraint [12]. The criticism is exactly the distinction drawn in Claim 7.3, between an artifact that carries evidence and one that carries a claim.

Auditor independence operates through a standard rather than through an occasion or a capacity, and it separates two things this paper has kept together. Professional frameworks require independence in fact together with independence in appearance: the first is a state of mind permitting a conclusion without being affected by influences that compromise professional judgment, and the second is the avoidance of circumstances that would lead a reasonable and informed third party to conclude that objectivity had been compromised [13]. The requirement is treated as fundamental to public confidence [31], and the frameworks identify categories of threat, including self-interest, self-review, advocacy, familiarity, and intimidation, against which safeguards are to be applied.

The two limbs correspond to two of this paper's concerns and the correspondence is worth stating. Independence in fact is the undertaking whose credibility Section 5.1 found hard to evidence, since

it concerns grounds rather than conduct. Independence in appearance is a requirement about what a third party would conclude, which is a requirement on observables and therefore checkable. That professional practice separates them, and requires both, is evidence that the difficulty this paper identifies is recognized in the field and addressed by requiring the checkable limb alongside the one that matters.

8.3 A Disclosed Personal Instrument

The author has constructed an instrument of the kind this paper theorises. It is described here, evaluated by the criteria applied above, and reported as failing several of them. The interest this creates is declared in the front matter.

The instrument is a personal declaration of permanent independence from state authority. Its operative content is four abstentions: from holding a career civil-service post or political appointment in any state; from joining any state's armed forces, intelligence services, or national security apparatus; from exercising administrative, diplomatic, law-enforcement or other coercive public authority for any state; and from accepting any position requiring exclusive political allegiance to a single state. It expressly reserves international organization work, work with non-governmental and non-profit bodies, independent research, university and educational work, public policy research, and peace mediation. Publication was contemplated as a personal page together with a timestamped record, and in several languages.

Three features favour it. The abstentions are stated as conduct rather than as disposition, so that their breach is observable in the way Section 7.4 requires: holding a named post is a public fact. They are differentially costly in the sense of Section 5.3, since they foreclose the state-employment and security career paths, and the foreclosure is expensive in proportion to how available those paths were. And the reservation clause performs a function the paper has not previously named, since by stating what the declarant will continue to do it prevents the abstentions from being read as a withdrawal from all public activity, which would make the instrument a statement about inclination rather than about capture.

Four failures are reported.

The instrument occupies the corner Claim 7.2 identifies as uncredible. The rule is stated publicly and the determination of breach remains with the declarant, since no party is designated to decide whether an abstention has been violated, and none is under any obligation to look. By Claim 7.1 this is the case in which credibility is not produced, and the instrument as constituted is therefore a statement rather than a commitment device.

Publication supplies observability of the rule and not of the conduct. A timestamped record establishes what was undertaken and when, which addresses one deficit of Section 5.2 and leaves the other two untouched. Nothing about the instrument makes the declarant's subsequent conduct visible, and nothing about it survives the declarant.

The instrument produces no artifact meeting Claim 7.3. What it generates is a declaration, which is an utterance in the first register of Section 6 whose weight is that of the declarant's own position. Since the declarant's position is what the instrument exists to establish, the artifact certifies its author.

And nothing binds the parties who might capture the declarant. This is the feature that distinguishes the international civil service regime, and its absence here is not a defect of drafting but a limit of the private form: a private person cannot impose obligations on states.

The instrument therefore fails as a commitment device and does something else which is worth distinguishing. It fixes the content of an undertaking, dates it, and makes it public, which are the conditions under which a later determination by others becomes possible. In the terms of Section 7.1

Table 2: Four instruments of individual independence

Instrument	Trilemma corner	Determination of breach	Available to a stranger
International civil service	Delegated	The Organization, under staff regulations and rules	The Charter provision, the rule, the fact of the witnessed declaration, and the reciprocal obligation on states
Recusal	Capability removal, occasion by occasion	The court or body, on stated grounds	The fact and the grounds of withdrawal from a named matter
Blind trust	Capability removal	The trustee arrangement, by construction	The existence and terms of the trust, but not what the officeholder recalls placing in it
Auditor independence	Delegated, with a standard	Professional and regulatory bodies applying a threats-and-safeguards framework	Circumstances bearing on independence in appearance
Personal declaration	Bare declaration	The declarant	The content and date of the undertaking, and nothing about conduct under it

it is self-legislation and not self-constitutionalization, and its value lies in what could be built on it rather than in what it presently accomplishes.

8.4 Findings of the Comparison

Table 2 locates the four instruments.

Three findings follow.

The instruments differ in which corner they occupy and none occupies more than one, which is the comparative evidence for Claim 7.2. The evidence is weak in the way a four-case comparison is weak, and it would be defeated by a single instrument holding all three properties.

The international civil service is the only instrument that binds the parties who might capture the committed individual. The other three constrain the individual alone, and therefore address the problem from one side. This is the principal finding of the comparison and it identifies what a private instrument structurally cannot do, since obligations on states are not within a private person's power to create.

And the instruments that work best are the ones that require the least of anyone's state of mind. Recusal asks nothing about disposition; the blind trust asks nothing about resolve; Article 100 asks for a declaration and then relies on an employer to enforce it. The instrument examined in Section 8.3 asks the most about disposition and supplies the least by way of enforcement, which is the same finding stated from the other end.

9 Fragility, Entrapment, and Depletion

The preceding sections concerned how credibility is produced. This section concerns how it is lost, and identifies three exposures that attach to the individual position rather than to any particular instrument.

9.1 The Fragility Asymmetry of a Short Record

Reputation is a belief held by others about a party's type, revised as conduct is observed. The formal treatment of reputation in repeated interaction rests on exactly this structure, in which a small prior probability that a party is of a committed type sustains behaviour that would otherwise unravel, and in which observed conduct updates that probability [37, 38, 45].

Two consequences follow for a party whose record is short, and they are consequences of the updating structure rather than of anything about individuals.

A belief supported by few observations moves further when a new observation arrives. An institution with a long record of consistent conduct has a belief about it that a single inconsistent act shifts slightly; an individual with a short record has a belief that the same act may reverse. The individual's reputation is therefore more volatile in both directions, which makes it quicker to establish and quicker to destroy than the volatility of the belief alone would suggest.

Volatility is not symmetric in practice, and the second consequence follows from that. Slovic's account of trust in risk management states an asymmetry principle: trust-relevant favourable characteristics are judged to require many instances to establish and few to lose, and sources conveying trust-destroying information are judged more credible than sources conveying the reverse [58]. Combined with the short record, the effect compounds. An individual custodian's standing is built slowly by observations that each move the belief little in the favourable direction, and can be undone by a single observation that moves it far in the unfavourable one.

Claim 9.1 (Fragility asymmetry)**.** An individual custodian's reputation is both harder to establish and easier to destroy than that of an institution performing the same function, because the belief rests on fewer observations and because unfavourable observations are weighted more heavily than favourable ones.

Claim 9.1 is an inference from the updating structure together with the asymmetry finding, and is not a result stated in either literature. The design consequence is the part that matters. An instrument designed for maximum stringency maximizes the number of ways in which a single observation can be unfavourable, which is the wrong objective for a party whose belief is supported by few observations. The correct objective is robustness to single events: fewer undertakings, each stated in terms whose breach is unambiguous, with a procedure for contesting an alleged breach before the belief is revised. The instrument examined in Section 8.3 has the first two properties and lacks the third.

9.2 Entrapment and the Stringency Trap

A published undertaking supplies a party who wishes to damage the declarant with a specification of what would constitute damage. The more stringent and the more numerous the undertakings, the larger the set of circumstances that can be arranged or represented as a breach.

The exposure has two forms. In the first, a violation is manufactured: the declarant is placed in circumstances designed to produce conduct falling within the terms of an abstention, or appearing to. In the second, no arrangement is necessary and an ambiguous episode is characterized as a breach, which succeeds where the terms admit interpretation and where, by Claim 9.1, the unfavourable characterization is weighted heavily against a short record.

An institution is protected against both by properties an individual lacks. Its record absorbs a single episode; it has a procedure for contesting a characterization; and it can survive the interval during which a contested matter is resolved. A private declarant has no procedure, no interval, and no reserve of prior observations.

Two design consequences follow and they are in tension with the intuition that a stronger undertaking is a better one. Undertakings should be few and their terms unambiguous, since ambiguity is the surface on which the second form operates. And an instrument should specify in advance how an alleged breach is to be examined, which is a further reason for the delegated corner of Claim 7.2: the party that determines breach is also the party that can decline to find one.

The practice this describes is documented, though for a different class of target. Ledeneva's study of informal practices in post-Soviet politics examines *kompromat*, compromising material that may be collected, stored, traded, or deployed strategically against a public figure, and records that much of its force lies in its unpublished form, where the threat of release is used to alter conduct rather than to destroy standing [40]. The practice has received treatment in the literature on media and scandal [61]. Two features of it bear on the account. The material may be genuine or fabricated and operates either way, which is the second form of the exposure described above. And its unpublished use is coercive rather than destructive, which identifies a route to capture that the paper has not otherwise named: a custodian may be induced to concede by the prospect of a disclosure that has not occurred.

What the literature examines is politicians, officials, and business figures, and its object is standing in general rather than a published undertaking in particular. The application here, in which a declarant's own stated abstentions supply the specification of what would constitute a breach, is not found in that literature and is advanced on the argument given above. Section 11 records the claim as supported in its general form and unsupported in its specific one.

9.3 Capture of the Enforcement Community

Claim 7.1 holds that credibility requires transferring the determination of breach to others. Those others thereby acquire authority over the declarant, and nothing in the arrangement makes them incorruptible.

The exposure is structural rather than incidental. A party that wishes to capture a custodian has two routes: the custodian, and whoever determines whether the custodian has been captured. The second is frequently the cheaper, because a determining body is smaller, is often less visible than the custodian whose conduct it superintends, and holds an authority whose exercise is not itself superintended. The remedy for capture therefore reproduces the problem one level up, and it does so in a form that is harder to observe.

Three observations bound the exposure without removing it. A determining body whose findings are published exposes its own conduct to the same inference that the custodian's conduct is exposed to, which is one reason the artifact condition of Claim 7.3 matters at this level as well. A distributed determination, in which several parties must concur, raises the cost of capture in proportion to the number and independence of the parties, which is the structure the private-order communities exhibit [46, 28, 51]. And a determining body that has its own reputation at stake in the accuracy of its findings has an interest against capture, which is the mechanism by which certification bodies are supposed to function and which fails when the body's reputation is less valuable to it than the relationship it is asked to protect.

The exposure is not eliminable within the account. A regress of determining bodies terminates somewhere, and wherever it terminates the terminating party determines its own breach. What the account can say is that the regress should terminate as far from the interested parties as the arrangement permits, and that the termination point should be visible rather than implicit.

9.4 The Custodial Burden and Depletion

A final exposure concerns the cost of occupying the position rather than the loss of it.

A custodian is asked, repeatedly and by both parties, to demonstrate that it has not been captured. Each demand is particular, arrives at a moment chosen by the party making it, and is answered by an act rather than by a rule. Refusal is legible to the party refused and largely invisible to everyone else, so the refusals accumulate as costs without accumulating as a record. The benefit, meanwhile, is diffuse: what is preserved is the field, which no party holds and for which no party pays.

The structure of the burden differs between an office and a person, and the difference is the one identified in Section 2.2. A demand made on an office is answered by the office's rules, and the answering costs the officeholder nothing personally because the rule and not the person is what refuses. A demand made on a person is answered by the person, and is answered again at the next demand. This is a further respect in which Claim 7.2's delegated corner is favourable: a rule administered by others converts a relational question, which must be answered each time, into an institutional one, which is answered once.

The depletion is not merely a matter of effort. A custodian who is worn down either abandons the position or begins to answer demands by conceding to them, which is capture arriving through exhaustion rather than through inducement. The paper records this as an exposure of the individual position and does not develop it; a fuller treatment of the cost borne by parties who maintain relational conditions, and of why that cost is not met by those who benefit from it, is deferred to separate work.

10 Implications for the Generative Relational Framework

This section states what the analysis returns to the wider framework. It is a discussion section rather than a further exposition, and it records four results, two of which qualify positions the companion paper took.

10.1 Credibility among the Conditions of Custodianship

The companion paper treated a custodian as an actor that maintains the conditions constituting a field, and constrained the instruments it may use. It took for granted that a custodian is accepted as one. The present analysis shows that acceptance is itself a condition and one that must be produced, since a party whose undertaking of non-capture is not believed is not used, and a custodian that is not used maintains nothing.

Credibility therefore belongs among the conditions the framework treats as requiring governance rather than among the presuppositions of the analysis. This is a correction to the companion paper's scope: the maintenance of a field depends on a further condition, internal to the custodian's position, that the companion paper did not identify.

10.2 The Register of a Relational Condition

Section 6.5 recorded that relational conditions differ in the register through which they operate, and the consequence for the framework is worth restating here.

A field composed of conditions operating in a single register carries that register's characteristic failure as a systemic risk. A field constituted entirely by accredited venues and recognized offices fails when recognition weakens, and does so across all its conditions at once. A field constituted

entirely by histories of joint work is robust to that failure and cannot be extended to parties who share none of those histories. The design implication is that a field's resilience depends on the distribution of its conditions across registers and not only on their number, which is a consideration the companion paper's treatment of field maintenance did not contain.

10.3 The Cost of Revisability

The framework treats revisability as a requirement: no arrangement should become irreversible, and no interpretation should be placed beyond reopening. Claim 7.2 shows that this requirement is not free.

A commitment device that secures credibility while preserving revisability must cede the authority to determine what the commitment requires. A framework committed to revisability is therefore committed to the cession, and cannot consistently prefer arrangements in which an actor both binds itself credibly and retains interpretive authority over its own binding. The cession is an entailment of the framework's own commitments rather than a concession to circumstance, and it should be stated as such wherever the framework recommends revisable arrangements.

A second consequence bears on the framework's treatment of interpretive authority more generally. Wherever the framework holds that an interpretation must remain contestable, it thereby locates the authority to sustain or reject that interpretation somewhere other than with the party interpreted. The capture exposure of Section 9.3 attaches to that location, and the framework has no general account of how the party holding it is itself constrained.

10.4 Individual and Institutional Scale

The companion paper closed by observing that a state's custodianship is exercised through an office and an individual's through a person. The present analysis identifies what turns on that difference and what does not.

What does not turn on it: the constraint on field-constitutive instruments applies unchanged at both scales, and the registers of trust production operate at both. What turns on it: the properties supporting a belief in the undertaking, which an office holds by construction and a person must acquire; the volatility of that belief, which is a function of record length; and the structure of the burden, which an office answers by rule and a person answers by act.

A further scale is visible and untreated. An organisation that is neither a state nor a person occupies an intermediate position, holding duration and some observability while lacking the reciprocal obligations that bind states in respect of an international civil service. Whether the account developed here applies to such bodies without modification is not established, and Section 11 records it.

11 Limits of the Account

This section records where the account stops. Several limits were found in the course of the analysis rather than anticipated, and those are marked.

11.1 Conditions of Falsification

Section 4.3 stated five conditions. Their status after the analysis is as follows.

The claim that the three deficits attach to the individual position was qualified during the analysis rather than after it. Section 5.2 concedes that a community supplies duration, observability

and exit cost to its members, and retains only the residue that a community supplies these in respect of conduct it observes and cares about, which may not include capture by an external party. The residue is argued and not evidenced.

The trilemma is supported by a four-instrument comparison in which no instrument occupies more than one corner. That is weak evidence of the kind a four-case comparison provides, and it would be defeated by a single instrument holding credibility, revisability and retained authority together.

The registers were not shown to be independent. Section 6 argues that they have distinct requirements and distinct failures, and supplies no case in which one was secured and the others demonstrably were not. The typology is therefore a proposal supported by argument.

The fragility asymmetry is an inference from two literatures rather than a result stated in either [37, 58], as Section 9.1 records. No measurement of the relative volatility of individual and institutional reputations is offered.

The condition that the paper's prescriptions might worsen conduct is partly satisfied. Section 7.4 retracts the artifact prescription in its original form on the strength of evidence that disclosure increases the bias it reports [10, 11], and Claim 7.3 restates it narrowly. Whether the narrowed form escapes the finding is not established, since the studies concern disclosure of interest rather than artifacts of the kind the claim requires.

11.2 Claims Advanced Without Support

Three claims are advanced on argument alone and are marked as such.

The demand-side conjecture of Section 5.4, that non-alignment is a positive signal in one class of market and a negative signal in another, was not supported by any empirical literature located in the survey underlying this paper. It is testable and untested.

The entrapment exposure of Section 9.2 likewise rests on argument. No scholarly treatment of the deliberate manufacture of a reputational violation against a custodial position was located.

And the residue argument of Section 5.2, that a community does not observe or penalise capture by an external party, is a conjecture about what communities attend to, on which no evidence is offered.

11.3 Sources Not Yet Verified

This draft cites only sources verified against a publisher, journal, governmental or institutional page. Several literatures the argument would ordinarily draw on are consequently absent, and the absences are substantive rather than stylistic.

The diachronic-agency literature is represented only in part. Section 3.5 draws on the accounts of planning agency and of resolution [8, 9, 30] and on reductionism about personal identity [52], and omits the literature on identification and on the sourcehood conditions of responsibility, which bears on whether the undertaking is the declarant's own in the sense that would make its abandonment a betrayal rather than a change of mind. Schelling on commitment and self-command, Fearon on audience costs, and the bonding-cost literature are absent from Section 3.2, which is accordingly narrower than the field it reports. Holmes on constitutions as enabling is absent from Section 3.3. The accounts of trust as reliance on goodwill are now present [4, 33, 34, 29] and Section 3.1 states the objection they support, which this paper does not answer: that credibility engineered by design is not trust, and that an account of how commitments are made believable is available to a party who intends to defect.

11.4 Extensions

Four extensions are identified and none is attempted.

An account of intermediate bodies, neither states nor persons, would establish whether the analysis transfers to organisations, which is the case most relevant to practice.

An empirical treatment of the demand-side conjecture would convert Section 5.4 from a conjecture into a finding or dispose of it.

A treatment of the cost borne by parties who maintain relational conditions, and of why that cost is not met by those who benefit from it, is required by Section 9.4 and deferred from it.

And a design study of instruments occupying the delegated corner, addressing how a determining body is constituted, how its findings are published, and how its own capture is guarded against, is what Section 9.3 shows to be necessary and this paper does not supply.

12 Conclusion

A state and a person may give the same undertaking and be believed unequally, for reasons that have nothing to do with either one's sincerity. The state will outlive the undertaking, its conduct is recorded, and defection would destroy arrangements that others rely on. This paper began from the observation that an individual holds none of these, and asked what follows.

What follows is that the devices by which individuals make commitments believable all operate in one way. They transfer the determination of breach, in whole or in part, to parties other than the declarant, either by relocating the judgment or by removing the occasion for it. There is accordingly no purely self-imposed credibility, and self-legislation names the origin of a rule rather than the arrangement by which it holds. The general point is not new, and the paper concedes it to Elster and to the literature on enforcement without a sovereign; what is added is the individual case, in which the undertaking concerns conduct toward parties outside any community that observes the declarant.

From this a constraint follows. Credibility, revisability, and retained authority over the determination of one's own breach are jointly unavailable. Removing the capacity to defect secures the first two of these and forecloses revision; declaring a rule one judges oneself preserves revision and authority and persuades nobody; delegating the determination secures credibility and revision and cedes the authority. A framework that requires revisability is thereby committed to ceding authority, which is an entailment rather than a regrettable cost.

The paper further distinguishes three registers through which trust is produced, and finds that they substitute poorly for one another. Positional guarantee is fast and transferable and fails all at once when the guaranteeing order weakens. What a shared ordeal establishes is strong and cannot be transferred to a party who was not present, which is why individual credibility does not accumulate with experience the way competence does. And trust by identification is structurally adverse to a custodian, since succeeding at it with one party supplies the other with a reason for distrust.

Four instruments were compared. Only the international civil service binds the parties who might capture the committed individual, and it does so by a provision no private instrument can replicate. The instruments that work best require the least of anyone's state of mind. And the instrument the author constructed, examined here and disclosed as an interest, fails as a commitment device: it states a rule publicly and leaves the determination of breach with its author, which is the corner in which credibility is not produced. What it achieves is to fix the content of an undertaking and date it, which are the conditions under which a determination by others could later be built.

Two exposures attach to the position rather than to any instrument. A short record makes an individual's standing both slower to establish and quicker to lose, which argues for fewer and clearer undertakings rather than more stringent ones, and which leaves the declarant exposed to a violation that another party arranges or alleges. And the community recruited to determine breach is itself capturable, so that the remedy reproduces the problem at a level that is harder to observe and cannot be regressed away.

The account is a revisable proposal. It supplies a diagnosis, a typology, a derived constraint, and a comparison; it supplies no measurement of the volatility it asserts, no evidence for the market conjecture it advances, and no design for the determining body its own conclusion requires. It also retracts one of its own prescriptions on the strength of evidence that disclosure can worsen what it reports. Those are the next requirements, and the account should be judged by whether they can be met.

Acknowledgments

The present definitions, constructions, arguments, conclusions, and errors remain the author's responsibility. The instrument examined in Section 8.3 was constructed by the author, and the interest this creates is declared in the front matter.

References

[1] Mohammad Akbarpour and Shengwu Li. Credible auctions: A trilemma. *Econometrica*, 88(2):425–467, 2020.

[2] Arthur Isak Applbaum. *Ethics for Adversaries: The Morality of Roles in Public and Professional Life*. Princeton University Press, 1999.

[3] Lewis Aron. *A Meeting of Minds: Mutuality in Psychoanalysis*. The Analytic Press, 1996.

[4] Annette Baier. Trust and antitrust. *Ethics*, 96(2):231–260, 1986.

[5] Jessica Benjamin. Beyond doer and done to: An intersubjective view of thirdness. *The Psychoanalytic Quarterly*, 73(1):5–46, 2004.

[6] Joyce Berg, John Dickhaut, and Kevin McCabe. Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1):122–142, 1995.

[7] Lisa Bernstein. Opting out of the legal system: Extralegal contractual relations in the diamond industry. *Journal of Legal Studies*, 21(1):115–157, 1992.

[8] Michael E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA, 1987.

[9] Michael E. Bratman. *Structures of Agency*. Oxford University Press, Oxford, 2007.

[10] Daylian M. Cain, George Loewenstein, and Don A. Moore. The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *Journal of Legal Studies*, 34(1):1–25, 2005.

[11] Daylian M. Cain, George Loewenstein, and Don A. Moore. When sunlight fails to disinfect: Understanding the perverse effects of disclosing conflicts of interest. *Journal of Consumer Research*, 37(5), 2011.

[12] CBC News. Carney set up a blind trust, screens to avoid conflicts of interest. is that enough? CBC News. https://www.cbc.ca/news/politics/carney-conflict-of-interest-brookfield-blind-trust-1.7493849.

[13] Professional independence resources. *The CPA Journal*, 2024. Summarising the AICPA definitions of independence of mind and independence in appearance. https://www.cpajournal.com/2024/04/30/professional-independence-resources/.

[14] Jeremy Edwards and Sheilagh Ogilvie. Contract enforcement, institutions, and social capital: The Maghribi traders reappraised. *Economic History Review*, 65(2):421–444, 2012.

[15] Robert C. Ellickson. *Order Without Law: How Neighbors Settle Disputes*. Harvard University Press, 1991.

[16] Jon Elster. *Ulysses and the Sirens: Studies in Rationality and Irrationality*. Cambridge University Press, 1979.

[17] Jon Elster. *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*. Cambridge University Press, 2000.

[18] Mustafa Emirbayer. Manifesto for a relational sociology. *American Journal of Sociology*, 103(2):281–317, 1997.

[19] Mustafa Emirbayer and Ann Mische. What is agency? *American Journal of Sociology*, 103(4):962–1023, 1998.

[20] Erik H. Erikson. *Childhood and Society*. W. W. Norton, 1950.

[21] Bruce Fink. *The Lacanian Subject: Between Language and Jouissance*. Princeton University Press, 1995.

[22] Diego Gambetta. *Codes of the Underworld: How Criminals Communicate*. Princeton University Press, 2009. ISBN 9780691152479; JSTOR j.ctt7tbn3.

[23] Diego Gambetta and Heather Hamill. *Streetwise: How Taxi Drivers Establish Their Customers' Trustworthiness*. Russell Sage Foundation, 2005.

[24] Kenneth J. Gergen. *Relational Being: Beyond Self and Community*. Oxford University Press, 2009.

[25] Peter Goodrich. *Oedipus Lex: Psychoanalysis, History, Law*. University of California Press, 1995.

[26] Jay Greenberg. The problem of analytic neutrality. *Contemporary Psychoanalysis*, 22(1):76–86, 1986.

[27] Avner Greif. Reputation and coalitions in medieval trade: Evidence on the Maghribi traders. *Journal of Economic History*, 49(4):857–882, 1989.

[28] Avner Greif. Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition. *American Economic Review*, 83(3):525–548, 1993.

[29] Katherine Hawley. Trust, distrust and commitment. *Noûs*, 48(1):1–20, 2014.

[30] Richard Holton. *Willing, Wanting, Waiting*. Clarendon Press, Oxford University Press, 2009. 203 pp. ISBN 9780199214570.

[31] International Ethics Standards Board for Accountants. IESBA addresses independence expectations in group audits. Ethics Board, 2022. `https://www.ethicsboard.org/news-events/2022-02/iesba-addresses-independence-expectations-group-audits`.

[32] Noel D. Johnson and Alexandra A. Mislin. Trust games: A meta-analysis. *Journal of Economic Psychology*, 32(5):865–889, 2011.

[33] Karen Jones. Trust as an affective attitude. *Ethics*, 107(1):4–25, 1996.

[34] Karen Jones. Trustworthiness. *Ethics*, 123(1):61–85, 2012.

[35] Christine M. Korsgaard. *The Sources of Normativity*. Cambridge University Press, 1996.

[36] Christine M. Korsgaard. *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press, 2009.

[37] David M. Kreps, Paul Milgrom, John Roberts, and Robert Wilson. Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, 27(2):245–252, 1982.

[38] David M. Kreps and Robert Wilson. Reputation and imperfect information. *Journal of Economic Theory*, 27(2):253–279, 1982.

[39] Jacques Lacan. The function and field of speech and language in psychoanalysis. In *Écrits: The First Complete Edition in English*, pages 197–268. W. W. Norton, 2006. Originally delivered 1953. Page range per the Norton complete edition; some sources give the French 1966 pagination.

[40] Alena V. Ledeneva. *How Russia Really Works: The Informal Practices That Shaped Post-Soviet Politics and Business*. Cornell University Press, 2006.

[41] Niklas Luhmann. *Trust and Power*. John Wiley, 1979. Expanded edition, Polity Press, 2017.

[42] Niklas Luhmann. Familiarity, confidence, trust: Problems and alternatives. In Diego Gambetta, editor, *Trust: Making and Breaking Cooperative Relations*, pages 94–107. Blackwell, 1988.

[43] Roger C. Mayer, James H. Davis, and F. David Schoorman. An integrative model of organizational trust. *Academy of Management Review*, 20(3):709–734, 1995.

[44] Debra Meyerson, Karl E. Weick, and Roderick M. Kramer. Swift trust and temporary groups. In Roderick M. Kramer and Tom R. Tyler, editors, *Trust in Organizations: Frontiers of Theory and Research*, pages 166–195. SAGE, 1996.

[45] Paul Milgrom and John Roberts. Predation, reputation, and entry deterrence. *Journal of Economic Theory*, 27(2):280–312, 1982.

[46] Paul R. Milgrom, Douglass C. North, and Barry R. Weingast. The role of institutions in the revival of trade: The medieval law merchant, private judges, and the Champagne fairs. *Economics and Politics*, 2(1):1–23, 1990.

[47] Stephen A. Mitchell. *Relational Concepts in Psychoanalysis: An Integration*. Harvard University Press, 1988.

[48] Edward Nersessian. Neutrality and curiosity. *The Psychoanalytic Quarterly*, 76, 2007.

[49] Andreas Ortmann, John Fitzgerald, and Carl Boeing. Trust, reciprocity, and social history: A re-examination. *Experimental Economics*, 3(1):81–100, 2000.

[50] Tom O'Shea. A law of one's own: Self-legislation and radical Kantian constructivism. *European Journal of Philosophy*, 23(4), 2015.

[51] Elinor Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, 1990.

[52] Derek Parfit. *Reasons and Persons*. Oxford University Press, Oxford, 1984.

[53] RAND Corporation. Addressing common questions and concerns from dual or naturalized citizens regarding the security clearance process. Technical Report RR-A3461-2, RAND Corporation, 2025. On the National Security Adjudicative Guidelines of Security Executive Agent Directive 4. `https://www.rand.org/pubs/research_reports/RRA3461-2.html`.

[54] Denise M. Rousseau, Sim B. Sitkin, Ronald S. Burt, and Colin Camerer. Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3):393–404, 1998.

[55] Renata Salecl. *The Spoils of Freedom: Psychoanalysis and Feminism after the Fall of Socialism*. Routledge, 1994.

[56] David Schrans. The individual and the collective: Sociological influences on Lacan's concept of the relation subject–other. *Frontiers in Psychology*, 9:614, 2018.

[57] Bradley N. Seeman. The "paradox of self-constitution" and Korsgaard's two conceptions of maxims for action. *Metaphilosophy*, 47(2):233–250, 2016.

[58] Paul Slovic. Perceived risk, trust, and democracy. *Risk Analysis*, 13(6):675–682, 1993.

[59] Michael Spence. Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374, 1973.

[60] Cass R. Sunstein and Edna Ullmann-Margalit. Second-order decisions. *Ethics*, 110(1):5–31, 1999.

[61] Howard Tumber and Silvio Waisbord, editors. *The Routledge Companion to Media and Scandal*. Routledge. Includes a chapter on political scandal and kompromat. ISBN 9781032093192.

[62] Charter of the united nations, article 100. Repertory of Practice of United Nations Organs, Codification Division, United Nations. `https://legal.un.org/repertory/art100.shtml`.

[63] United Nations. Staff regulations and rules: Duties, obligations and privileges, article i, regulation 1.1. United Nations Policy Portal. `https://policy.un.org/node/32`.

[64] United Nations Secretary-General. Status, basic rights and duties of united nations staff members. Secretary-General's Bulletin ST/SGB/2003/5, 2003. `https://digitallibrary.un.org/record/488771/files/ST_SGB_2003_5-EN.pdf`.

[65] United States Department of State. Security clearance with dual citizenship. Bureau of Diplomatic Security, careers.state.gov. `https://careers.state.gov/uploads/c9/f9/c9f9652a5b1bea0f14295e4553020380/3.1.7_FSO_Security_Clearance_w_Dual_Citizenship.pdf`.

[66] Slavoj Žižek. *The Sublime Object of Ideology*. Verso, 1989.

[67] Slavoj Žižek. *The Ticklish Subject: The Absent Centre of Political Ontology*. Verso, 1999.

[68] Oliver E. Williamson. Credible commitments: Using hostages to support exchange. *American Economic Review*, 73(4):519–540, 1983. JSTOR 1816557.

[69] Lynne G. Zucker. Production of trust: Institutional sources of economic structure, 1840–1920. *Research in Organizational Behavior*, 8:53–111, 1986.