

Multiple Reference and the Reconstruction of Co-Experience: A Generative Relational Account of Recommendation Procedure

Wanhong HUANG
huangwanhong@serendip.ngo

Abstract

Institutions that select people by reference almost always ask for more than one. This paper takes that requirement seriously and asks what it is for. On the account developed here, the shared experience between a candidate and a referee is not an object that either party holds and one of them writes down. It is generated within their relation, and each account of it is a projection taken from one relational position. Asking for several references is therefore an attempt to reconstruct an object from its presentations, in the way that what identifies an attractor is what remains invariant across trajectories. The institution is doing something structurally correct and doing it badly. Four consequences follow for the design of the procedure. What matters among sources is independence of relational position rather than the number of letters. What carries information is what survives across projections rather than their average or their maximum. Symbolic authority accrues to those whose time is scarce, so an eminent referee supplies a thin projection, and weighting by authority is accordingly the least suitable weighting for a reconstruction. And the calibration of individual referees, which no institution maintains, is a precondition for correcting the distortion each introduces. Two failures of procedure are then identified: evaluators collapse onto the single most authoritative account, which discards the reconstruction and returns the decision to one projection; and the candidate, who is one of the two producers of the experience, contributes no account at all. The paper proposes that the candidate's structured account be collected as a required source at the time references are requested. It closes with the injustice that survives every correction it proposes, including a paradox in which contesting an account may destroy the relation that generated the standing to contest it. The paper's central empirical premise, that divergence among sources carries information, is contested in the literature it draws on, and the paper argues for it rather than assuming it.

Keywords: recommendation; selection procedure; multi-source assessment; rater divergence; co-experience; commensuration.

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 proposals made here. The paper recommends collecting more information about candidates and about referees than institutions presently collect, and recommendations of that kind can be implemented in ways that increase surveillance of the people they were intended to serve. The design criteria in this paper are stated together with the limits that govern them, and the author asks that the two be applied together. 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. Several literatures the argument bears on are represented only in part, as the accompanying literature audit records. Empirical evaluation of the procedural proposals, specialist review of the measurement material, 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 L^AT_EX. Every source cited here was verified before it was written into the manuscript rather than after. 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. The author is an applicant in processes that use the procedures examined here, and has been a subject of the instrument this paper analyses. The paper is written as a design analysis rather than as an account of any particular case, no individual process or institution is described, and the proposals are assessed by criteria stated in advance of them. The interest is declared because a paper recommending changes to a procedure the author is subject to should say so.

Related research programme. This paper is project P003 and the third 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; project P002 for the individual-scale credibility problem, the registers of trust production, and the commitment trilemma. The present paper takes responsibility for the reconstruction account of multiple reference, the procedural diagnosis, and the design criteria that follow. Later papers in the series take the ontology of the judged subject, the ethics of the referee, the distribution of generative conditions, and the jurisprudence of interpretive authority over shared experience; those questions are marked where they arise and are not argued here.

Suggested citation. Huang, Wanhong. “Multiple Reference and the Reconstruction of Co-Experience: A Generative Relational Account of Recommendation Procedure.” Working discussion paper, 2026.

1 Introduction

A graduate programme asks for three letters. A hiring committee asks for two referees and calls both. A fellowship asks for four, from people who have known the candidate in different capacities. The requirement is so ordinary that its content is rarely examined, and the examination is usually critical when it occurs: letters are shown to predict poorly, to be written in a register that inflates everything, and to reward candidates whose acquaintances are eminent.

This paper begins from a different observation. The requirement is for *more than one*, and that is a strange thing to ask for if the object is to obtain an opinion. One opinion would do, and a better-informed opinion would do better. Asking for several from people differently placed is what one does when no single account is expected to be sufficient, and when the several accounts are expected to bear on one another. Institutions that do this are attempting something more demanding than collecting opinions, and the paper takes the attempt seriously.

What they are attempting can be stated once the object is described correctly. The shared experience between a candidate and a referee is not a thing that either party holds and one of them writes down. It is generated within the relation between them, and it is available to each only from the position that party occupies. A supervisor sees what a supervisor is placed to see; a collaborator sees something else; a junior colleague, something else again. None of these is the experience, and none is a corrupted copy of it. Each is a projection, and the object stands behind them in the way that an unnamed regularity stands behind the observations from which it is eventually recognised. What identifies an attractor in a dynamical system is what remains invariant across its trajectories; what identifies an optimum in an economy is what remains invariant across allocations. Neither is read off a single instance.

If that is right, then requiring several references is an attempt at reconstruction, and the institution is doing something structurally correct. It is also doing it badly, and the paper's contribution is to say in what respects and what would follow from correcting them.

Four design consequences follow from the account and are developed in Section 6 and Section 7. What matters among sources is independence of relational position and not the number of letters, since three accounts from one laboratory are one projection taken three times. What carries information is what survives across projections, so that averaging them or taking the most favourable destroys the quantity that identifies the object. Symbolic authority accrues to those whose time is scarce, and scarce time means less contact with any particular candidate, so an eminent referee supplies a thin projection and weighting by authority is the least suitable weighting available for a reconstruction. And the distortion each referee introduces cannot be corrected without a record of how that referee's past accounts have stood up, which no institution maintains.

Two failures of procedure are then identified, and they are failures of what is done with the material rather than of what is collected. The first is collapse: several accounts are gathered, and the decision is taken on the one signed by the most eminent name, which discards the reconstruction and returns the matter to a single projection, and by the argument above to the thinnest one. The second is an omission. The candidate is one of the two parties in whose relation the experience was generated, and holds a projection from a position no other party occupies. In current practice the candidate supplies none. The paper proposes that a structured account by the candidate be collected as a required source, at the time references are requested and before the letters arrive, which places it beyond tailoring and supplies the conditions against which the other accounts are read.

The paper is constructive in posture and it continues past its proposals. Section 11 states the injustice that survives every correction proposed here. Positions of observation are unequally available, so the quality of a reconstruction tracks how relationally wealthy a candidate has been.

The co-producer contributes a projection and still holds no authority over the interpretation built from it. Some of what a relation generated is visible from no institutional position at all, and an instrument that improves at reconstructing what it can see grows more confident about what it cannot. And a candidate who contests an account risks the relation that generated the standing from which the contest is made, which is a cost the paper records and does not dissolve.

One commitment should be stated at the outset. The paper’s central empirical premise is that divergence among sources carries information about the object rather than about the sources alone. That premise is contested in the literature the paper draws on, where a substantial position treats between-source variance as method bias to be averaged away. Section 3.2 sets out both positions and Section 6 argues for the first rather than assuming it. A reader who is persuaded of the second will find that the paper’s diagnosis of collapse survives and its reconstruction account does not.

Section 2 supplies the practice and the evidence on its performance. Section 3 locates the account among the literatures on reference validity, multi-source assessment, reconstruction from partial views, institutional persistence, and commensuration. Section 4 states the method and the conditions of disconfirmation. Section 5 develops the generative relational account and the criteria it imposes. Section 6 states the reconstruction reading. Section 7 identifies where the instrument departs from it, and Section 8 treats the position of the referee. Section 9 treats the delegation of judgment. Section 10 sets out the corrections and Section 11 what survives them. Section 12 states what the analysis returns to the wider framework, Section 13 records the limits, and Section 14 consolidates the position.

2 Background and Preliminaries

This section describes the practice, reports what is known about how well it performs, and states the vocabulary the argument uses. Analysis is reserved for Section 5 onward.

2.1 Recommendation Practice and the Requirement of Multiple References

The practice examined here has four features, and each bears on the argument.

References are solicited from more than one person. The number varies by setting and the requirement itself is near-universal in academic admissions, academic appointment, and much professional hiring. The paper takes this requirement as its object rather than as a background detail.

The referees are ordinarily chosen by the candidate, from among people who have stood in some working relation to the candidate, and are ordinarily expected to differ in the capacity in which they knew the candidate. Institutions frequently express a preference for referees who observed different kinds of work.

The account solicited is ordinarily a free-form letter rather than a structured instrument. The alternative exists and performs differently: structured reference-check procedures, in which the referee answers fixed questions with scored responses, have been shown to reach acceptable reliability and criterion validity [53]. Where the paper speaks of the instrument’s failings, the free-form letter is meant.

And the account is ordinarily confidential to the candidate. In United States practice the legal position is explicit: the Family Educational Rights and Privacy Act provides a general right of a student to inspect education records, and its regulations permit that right to be *waived* in respect of confidential recommendations submitted for admission, employment, or the receipt of an honour [21]. Section 11 returns to the waiver. What matters here is that the material on which a decision turns is routinely unavailable to the person it concerns.

2.2 Predictive Validity of Reference-Based Selection

Three findings about the performance of this practice are established well enough to be treated as given, and a fourth is a caution about how the first three are usually quoted.

Weak prediction of later outcomes. The meta-analysis of letters of recommendation in college and graduate admissions finds them positively but weakly related to later outcomes, including grade point average, performance ratings, degree attainment, and research productivity, with the strongest incremental contribution appearing for degree attainment [32]. The authors read this as grounds for qualified hope rather than for dismissal. The estimates rest on few studies per cell, so the intervals are wide, and this paper accordingly treats the direction of the finding as established and the magnitude as uncertain.

Agreement between referees against agreement within a referee. This is the finding on which the paper's argument turns and it deserves stating carefully. Reference reliability has been reported at approximately .22, against approximately .50 for two supervisors rating the same employee [1]. More pointedly, the literature has long recorded that there is more agreement between two recommendations written by one person for two different applicants than between two people writing recommendations for the same person [2]. Section 6 argues that this is what one should expect of projections taken from different positions, and Section 3.2 records that the same result admits a second reading, on which the between-referee variance is error.

Near-universal favourability of accounts. Of a corpus of nearly seven thousand reference ratings, ninety-six per cent rated candidates above average and fewer than one per cent rated them below average or poor [1]. A signal that is sent at nearly full strength for nearly every candidate discriminates weakly by construction.

Caution on comparative validity coefficients. Comparative rankings of selection methods are widely quoted from a synthesis of eighty-five years of findings [49], in which reference checks appear well below cognitive and structured-interview measures. Those figures should now be quoted with care. A systematic re-examination of how range-restriction corrections have been constructed and applied in personnel selection meta-analyses concludes that the common approaches produce substantial overcorrection, and that the validity of many selection procedures for predicting job performance has therefore been substantially overestimated; revised estimates are offered [47]. This paper accordingly avoids resting any argument on a particular validity coefficient. Its use of this literature is confined to the three findings above, which concern reliability, range restriction, and the relative weakness of the letter, and which the correction does not disturb.

The conjunction the practice presents. An instrument of weak and uncertain predictive validity, whose sources agree with one another less than they agree with themselves, which is almost uniformly favourable, and whose content is withheld from its subject, is nonetheless required near-universally. Section 4.1 takes this conjunction as the paper's starting problem, and Section 3.6 supplies the standard explanation of why practices of this kind persist.

2.3 Vocabulary Carried from the Preceding Papers

The account uses a small vocabulary established in the companion papers, restated here at the length required to follow the argument, together with three terms this paper adds.

A *relation* is an ongoing process between parties rather than a state obtaining at a moment, and is described by what it produces.

The *generativity* of a relation is its capacity to continue producing such outcomes, including outcomes no party can specify in advance. Generativity is a property of the relation and not of either party.

A *relational condition* is an arrangement whose presence or absence changes which relations can be formed or continued, and the *field* of a set of parties is the set of arrangements consistent with those conditions.

Revisability is the requirement that an interpretation remain open to being reopened, and that no interpretation be placed beyond the reach of further interpretation.

Three terms are introduced here. *Co-experience* is what a relation generates between its parties: the undertakings, difficulties, judgments and understandings that arose in it and that belong to neither party alone. A *relational position* is the standpoint from which a party stands in a relation, which fixes what of the relation is available to that party. And a *projection* is an account of a co-experience given from one relational position. The three terms are given their work in Section 5 and Section 6.

The vocabulary belongs to a wider framework of generative relational analysis, which the present paper uses in the compact form stated in Section 5.1 and does not otherwise import.

3 Literature Review

This section locates the account among the literatures it depends on. Its organisation follows the argument rather than the practice: the first subsection treats the question on which the paper's central premise turns, and the remainder treat reconstruction, the ordering of evaluation, the subject's own account, institutional persistence, the portability of records, and delegated judgment. The findings on reference validity were given in Section 2.2 and are taken as given here.

One subsection reports a genuine division in the field. The paper's premise is that divergence among sources carries information about the object; a substantial position holds that it is measurement error. Both are set out, and Section 6 argues for the first.

3.1 Transfer of Trust between Parties

Before the question of divergence arises, there is a prior one: why an account of a relation is solicited at all, rather than the deciding party forming its own view.

The framing this paper adopts comes from an account of how trust was produced in an economy across a period of rapid migration and institutional change. Three modes are distinguished. *Process-based* trust rests on a record of exchange between the parties themselves. *Characteristic-based* trust rests on shared background, such as common origin or membership. *Institution-based* trust rests on formal structures external to both parties, including professional certification, intermediaries, and regulation; and the historical argument is that mobility eroded the first two and forced reliance on the third [63].

A selecting institution meeting a candidate for the first time has neither of the first two. It has no record of exchange with the candidate, and it is frequently selecting precisely across the boundaries that characteristic-based trust depends on. The third mode is what remains, and a reference is an instrument of it: an account supplied by a party the institution can locate, and weighed by what that party's position in a structure is worth.

Two features of that mode explain what the instrument does and what it cannot do. What is transferred is not the trust that existed in the original relation, since that rested on a history

the receiving institution was not party to; what is transferred is an inscription that the receiving structure recognises. And the transfer requires the receiving institution to trust the referee, which is a separate matter from whether the referee’s account is accurate. The distinction between trusting a person and having confidence in a system is drawn directly in the sociology of trust [34], and the difference matters here because an institution that has confidence in a system of references need form no view about any particular referee.

Two adjacent literatures bound what may be expected of a transferred account. Where parties must act together without a shared history, they proceed on expectations imported from roles and categories, act as though trust were present, and calibrate afterwards [39]. And where institutions are unavailable and every party has reason to misrepresent, what conveys trustworthiness is a signal costly enough that the party who would misrepresent would not send it [27] — the general condition being that a signal separates types only where it is sufficiently more costly for the type that would misrepresent itself [51]. Neither condition is well satisfied by a reference letter, which is cheap to write and, as Section 2.2 recorded, favourable in almost every instance.

This section supplies the problem that the remainder of the paper addresses. Institution-based trust is what a selecting institution must rely on, a reference is the instrument through which it is supplied, and the instrument transfers an inscription rather than the relation that produced it.

3.2 Multi-Source Assessment and Rater Disagreement

Where several people rate the same person, their ratings differ, and the interpretation of that difference is the question.

Decomposition of variance in multisource ratings. The most-cited variance decomposition analysed two large samples of managers, of two thousand three hundred and fifty and two thousand one hundred and forty-two, each rated by multiple supervisors, peers, and subordinates. It reported idiosyncratic rater effects of sixty-two and fifty-three per cent across the two datasets, combined ratee-performance effects of twenty-one and twenty-five per cent, and random error of eleven and eighteen per cent [50]. The dominant term is therefore attached to the individual rater rather than to the person rated, and the paper states this plainly because it cuts against the reading it will defend as much as for it.

Divergence read as difference of standpoint. A reanalysis in the same tradition recovers systematic effects associated with the rater’s *source* rather than with the rater as an individual, averaging approximately eight per cent of variance across two samples, and defends those source factors as genuine differences of perspective rather than as artifacts of method [30]. Related meta-analytic work reports that correlations between sources are low, with supervisor and peer ratings correlating at about .34, self and supervisor at about .22, and self and peer at about .19, while reliabilities within a source are markedly higher, at about .50 for supervisors, .37 for peers, and .30 for subordinates, and concludes that the sources hold somewhat different perspectives [9]. Peer and subordinate ratings have further been found to add incremental validity over supervisor ratings, which is to say that they carry information the supervisor’s rating does not [10].

Divergence read as measurement error. Against this, an extensive meta-analytic treatment argues that a general factor persists in performance ratings after halo and three further sources of measurement error are controlled, accounting at construct level for sixty per cent of total variance, and that construct-level correlations among rated dimensions are substantially inflated by halo, by thirty-three per cent for supervisory and sixty-three per cent for peer intrarater correlations [59]. On

this reading, much of what distinguishes one rater’s account from another’s is error, the appropriate response to which is aggregation, since averaging across raters is precisely what recovers a true score from parallel measurements. The reliability analyses in the same programme support that treatment [58]. A related assessment of why ratings track performance so weakly canvasses several explanations, including intentional distortion by raters, rather than treating the gap as perspective [41].

Consequence of the division for the present account. The two readings are not distinguished by the data alone, because the same variance component is named differently under each. What the perspective reading calls the standpoint from which a rater sees, the error reading calls the idiosyncrasy with which a rater rates, and no decomposition of numeric ratings settles which it is.

Three considerations bear on the choice and are developed in Section 6. The systematic component that the perspective reading recovers is small, at approximately eight per cent [30], and the paper reports it at that size. The evidence on both sides comes from numeric ratings of job performance on common dimensions, which is not the material at issue here, since a reference is a narrative account of a relation rather than a scaled rating of a trait, and it is at least open whether the two behave alike. And the decisive consideration is what the divergence is divergence *about*: aggregation is the correct treatment of parallel measurements of one quantity, and the paper’s account denies that the several referees are measuring one quantity in parallel.

Two further sources supply the apparatus for treating source variance as interpretable rather than as residue. The multitrait-multimethod matrix established that agreement between different methods measuring the same thing and disagreement between methods measuring different things are separately informative, and that method variance is a component to be identified rather than a nuisance to be removed [8]. Generalizability theory treats the rater as a facet of a measurement design whose variance is estimated as part of the analysis rather than assumed away [12]. Neither settles the question above, and both establish that treating between-source variance as an object of analysis is standard rather than eccentric.

3.3 Reconstruction of an Object from Partial Views

Three literatures bear on the claim that an object may be identified by what persists across its presentations.

Latent variables and the realism they require. Measurement models in psychology routinely posit an unobserved variable behind several observed indicators. The status of that variable has been argued directly, with the conclusion that a consistent interpretation of such models requires realism: the latent variable exists and stands in a causal relation to the indicators, rather than being a summary of them or a construction out of them [4]. The same programme develops the consequences for what validity is [5]. This supports the paper’s ontology at one point and constrains it at another. It supports treating an unobserved object as identified through its indicators. It constrains the paper by requiring a causal relation from object to indicator, and by warning that models estimated across persons do not license claims about the structure within a person.

Robustness and the independence of determinations. The philosophical statement closest to the paper’s use is that what can be detected, derived, or measured by several independent means is more trustworthy, and more likely to be real, than what rests on one [62]. The argument is explicitly about independence: the value of a second determination lies in its not sharing the failure

modes of the first. This is the ground of the paper’s requirement in Section 10.1 that what matters among sources is independence of position rather than number.

Attractor reconstruction and the limit of the analogy. The paper’s motivating image comes from dynamical systems, where an attractor is recovered from a single observed quantity by delay-coordinate embedding, the reconstruction being diffeomorphic to the original system and preserving its topological invariants [52]. The result carries conditions: the system must be deterministic, smooth, and finite-dimensional, the observable generic, and the embedding dimension sufficiently large relative to the dimension of the attractor. None of these conditions is satisfied by a relation between two people. The paper therefore uses the result as an analogy for what it means to identify something by its invariants and claims none of its guarantees, and Section 5.2 states this restriction where the analogy is used.

Triangulation for validation and triangulation for completeness. The methodological literature on combining sources supplies both the practice and its principal caution. The classical statement distinguishes triangulation of data, of investigators, of theories, and of methods [13]. The principal criticism holds that combining sources adds breadth rather than automatically conferring validation [22], and later work separates triangulation undertaken to validate a finding from triangulation undertaken to complete a picture [40, 23]. The distinction matters here and Section 6 states which is meant: the paper’s use is validation in respect of what is invariant, and completeness in respect of what is not.

3.4 Order and Contextual Bias in Evaluation

The paper proposes in Section 10.3 that accounts be read before their authors are known. Two literatures bear on whether such measures work, and their combined verdict is mixed.

Forensic science has developed the most explicit version of the procedure. After contextual bias was placed on the reform agenda by a national review of the field [43], a protocol was proposed under which the examiner analyses the trace evidence before being exposed to reference material or case context, with documented restrictions on revising the initial analysis once context arrives [18]. The approach has since been generalised beyond forensic comparison to expert decision-making more widely, with order effects documented directly [17]. The structure is exactly the one this paper proposes: fix the reading of the primary material before the information that would colour it becomes available.

The evidence on blinding in evaluation is more equivocal, and the paper reports it as such.

Blinding has been shown to change outcomes in the most-cited case, in which concealing the identity of auditioning musicians behind a screen increased by about half the probability that a woman would advance from certain preliminary rounds, and accounted for a substantial share of the subsequent rise in the proportion of women hired [28]. The authors themselves record that some estimates carry large standard errors and that one persistent effect runs in the opposite direction, and the causal estimate has been questioned since; the case is reported here with those qualifications.

In peer review, a randomised comparison of six hundred and seventy-four single-blind against seven hundred and eight double-blind manuscripts at one journal found that blinding author identities *lowered* ratings and acceptance rates, because it removed positive biases that had favoured authors from wealthy and English-speaking countries, while having no effect on gender differences in reviewer ratings [26]. Blinding therefore changes what is decided. Whether it improves the decision is a further question: an earlier randomised trial found that blinding made little or no difference

to the quality of reviews [57], while another found that blinded reviewers assessed prolific authors more evenly [37].

The lesson the paper takes from this is stated in Section 10.3. Removing an identity removes whatever that identity was carrying, which may be prejudice and may be information, and a procedure that removes it must say which it expects to remove and why.

3.5 The Subject’s Own Account as Evidence

The paper proposes in Section 10.2 that the candidate’s own structured account be collected as a required source. Two findings bound what may be expected of it.

Self-assessments agree poorly with the assessments of others. Self and supervisor ratings correlate at about .22 and self and peer ratings at about .19, which are among the lowest between-source correlations reported [9]. And the unstructured self-narrative performs weakly as a predictor: a meta-analysis of personal statements in admissions finds small relationships with later performance and little incremental validity over other measures [42].

A third finding separates the proposal from the personal statement. A structured self-report in which the candidate describes actual accomplishments against defined dimensions, and in which those descriptions are scored by trained raters rather than read impressionistically, has been shown to predict professional performance and to be largely uncorrelated with cognitive ability, which is to say that it carries information other instruments do not [31]. What is proposed in Section 10.2 is an instrument of this kind and not a statement of aspiration.

The low self-other agreement is read here as bearing on interpretation rather than as an objection. Under the account developed in Section 6, low agreement between the subject’s account and others’ accounts is what one expects of projections from different positions, and the question is what the divergence indicates rather than which party is correct.

3.6 Institutionalised Practice and Decoupling

The persistence of a weakly performing procedure has a standard explanation and the paper adopts it.

Formal structures are adopted in part because they conform to institutionalised accounts of what a proper organisation does, and their adoption confers legitimacy independently of whether they improve activity; organisations accordingly decouple their formal structures from their working practices and sustain the arrangement through a logic of confidence and good faith [38]. Organisations in a field come to resemble one another through coercive, mimetic, and normative pressures rather than through convergent discovery of what works [14].

Applied here, the near-universality of the requirement for two or three letters is what one expects of a practice that certifies the propriety of a selection process, and the weak relation between letters and outcomes is what one expects of a practice sustained on those grounds. The explanation and the paper’s account are compatible. Section 4.1 states the relation between them: the institutional account explains why a mis-specified instrument survives, and the paper’s account identifies what the instrument would be if it were specified correctly.

3.7 Portability of Inscriptions and Commensuration

A written account travels where a relation does not, and a literature on how records acquire that property bears directly on what a letter is.

The classical treatment concerns inscriptions that are at once mobile, in that they may be moved from place to place, and immutable, in that they do not deform under the movement; accumulating

such inscriptions at a centre permits actions there that are unavailable elsewhere [33]. The same treatment locates the achievement of linear perspective in its recognition of internal invariances under transformations produced by changes in spatial location, which is the property that makes a depiction transferable. Both elements bear on this paper: a reference is an immutable mobile, and what makes an account of a relation transferable is a question about invariance under change of standpoint.

Two further literatures describe the cost of that transfer. The pursuit of mechanical objectivity, in which judgment is displaced onto rules and numbers, is shown to arise where personal authority is weak and distrust is high, so that impersonality is a response to a political condition rather than a scientific advance [45]. And commensuration, the transformation of qualitative differences into quantities on a common scale, is analysed as a social process that renders some aspects of what it measures invisible or irrelevant, and as a form of power for that reason [20]. Contemporary work on classification extends this to the sorting of persons, in which ordinal position shapes life chances while presenting itself as merit [24, 25].

Section 6.5 draws the consequence the paper needs, that what makes an account portable is not the same as what makes it true of the relation it reports.

3.8 Delegated Judgment and the Loss of Capacity

An institution that relies on a referee’s judgment is depending epistemically on another party, and three literatures describe what that dependence costs.

The epistemological position is that a party may know something on the basis of another’s testimony without possessing the evidence, and that such dependence is rational rather than a failure, given the division of cognitive labour [29]. The dependence is therefore not objectionable in itself, which the paper accepts.

What follows from sustained dependence is a different matter. Automating a task removes the occasions on which the operator would have exercised the corresponding skill, so that the operator is least able to intervene at the moment when intervention is required [3]. Complacency and bias in the use of automated aids share an attentional basis, occur in experts as well as novices, and are not removed by training or by awareness [44]. The industrial statement of the general structure is the separation of conception from execution and the consequent loss of the capacity to conceive [6], though its author resisted the reading on which deskilling proceeds uniformly.

Section 9 argues that an institution which decides on referees’ judgments loses, over time, the capacity to form its own judgment of a person, and that this is a loss of capacity rather than of information.

3.9 Boundary of the Present Contribution

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

Four positions are left unoccupied by the literatures surveyed. No treatment located here reads the requirement of multiple references as an attempt to reconstruct an object from its presentations. None proposes that the subject’s own structured account be collected as a required source before the third-party accounts arrive. None proposes maintaining calibration records for individual referees. And none states the paradox developed in Section 11.4, in which contesting an account may destroy the relation that generated the standing from which the contest is made. The claim is that these four are unoccupied, not that their components are unprecedented; the components are conceded above.

Table 1: Antecedent literatures and contribution boundaries

Literature	Licensed role	P003 boundary
Reference validity	Weak prediction, low between-referee agreement, near-universal favourability [32, 1, 2]	Supplies the anomaly. It leaves open what the requirement of several references is for.
Multi-source assessment	Source and rater variance decomposed [50, 30, 9], against the reading of that variance as error [59]	The division is unresolved in that literature; this paper argues one side of it for narrative accounts of relations.
Reconstruction from partial views	Latent variables [4], robustness [62], invariants [52], triangulation [13]	Supplies the apparatus. Its transposition to co-experience is the present claim, and the dynamical result is used as analogy only.
Order and contextual bias	Sequencing protocols [18]; mixed evidence on blinding [28, 26, 57]	Establishes that ordering matters and that removing an identity removes information as well as prejudice.
Self-report instruments	Structured accomplishment records predict [31]; unstructured statements do not [42]	The proposal that the subject’s account be a required source among others is not found there.
Institutional persistence	Legitimacy, decoupling, isomorphism [38, 14]	Explains survival of the practice; does not identify what it is an attempt at.
Portability and commensuration	Immutable mobiles [33]; mechanical objectivity [45]; commensuration [20]	Describes what transfer costs; the reconstruction reading of why several transfers are demanded is the present contribution.
Delegated judgment	Epistemic dependence [29]; capacity loss under automation [3, 44]	Supplies the mechanism; its application to institutional knowledge of persons is developed here.

A terminological caution belongs here. Both *recommendation* and *calibration* carry established and entirely different meanings in research on algorithmic recommender systems, where a calibrated recommendation is one whose distribution of suggested items matches the distribution of a user’s past preferences. This paper uses recommendation for an account of a person given by a referee, and calibration in the forecasting sense of the correspondence between stated confidence and realised outcome [7]. No continuity with the recommender-systems literature is claimed or intended.

A bounded search for prior use of the paper’s own formulations, covering the reconstruction of co-experience, the account of a reference as a projection from a relational position, and the calibration of individual referees, returned no related scholarly use. A bounded search establishes that a formulation was not found rather than that it does not exist, and a systematic originality audit remains outstanding and is recorded in Section 13.

Four questions arising in this material belong to later papers in the series and are marked where they arise rather than argued here: what can in principle be known of a person from a relational position, which is the subject of the ontology paper; what a referee may assert given the access they had, which is the subject of the ethics paper; how unequal generative conditions produce unequal accounts, which is the subject of the injustice paper; and who holds authority to interpret a shared experience, which is the subject of the jurisprudence paper.

4 Method and Case Selection

This section states the form of explanation the paper attempts, what the comparative material is asked to do, and what would count against the account.

4.1 Anomaly-Driven Explanation

The paper’s method is to explain a practice by identifying what it is an attempt at, taking as its starting problem a conjunction that the practice’s stated purpose fails to explain.

The conjunction was set out in Section 2.2. An instrument of weak and uncertain predictive validity, whose sources agree with one another less than they agree with themselves, which is favourable in almost every instance, and whose content is withheld from its subject, is required almost universally in academic and professional selection. If the instrument were performing the function usually ascribed to it, that of informing a prediction about the candidate, its persistence would be difficult to account for.

Two explanations are available and they are compatible. The institutional account holds that formal structures persist because they confer legitimacy independently of their contribution to activity, and that organisations decouple such structures from their working practices [38, 14]. That account explains survival. It leaves open what the practice would be if it were performing well, and it therefore supports no proposal.

The account developed here asks a different question, namely what the requirement of *several* references is an attempt at, and answers that it is an attempt at reconstruction. That answer is assessed by whether it makes the features of the practice intelligible, whether it identifies departures that can be stated independently of it, and whether the corrections it implies are ones an institution could adopt. It is assessed independently of whether the practice performs well, which it fails to do.

The paper’s posture is accordingly constructive. It holds that the institution is attempting something structurally correct and executing it badly, and its contribution is to say in what respects, what would follow from correcting them, and what remains wrong when they have been corrected.

4.2 Selection of the Comparative Cases

Two comparative cases are used and each is chosen for a specific structural resemblance rather than for its subject matter.

Forensic examination under contextual bias is used in Section 10.3 because it is a setting in which an expert reading of primary material is known to be affected by information about its source, and in which a procedure has been developed to fix the reading before that information arrives [18, 17]. The resemblance is to the ordering problem and not to the subject matter.

Reliance on delegated assessment in financial markets is used in Section 9.3 because it is the largest documented instance of an institution’s assessment capacity atrophying as it came to rely on an external assessor, and because it has a reform history.

Two absences are recorded. Structured reference-check procedures, which perform differently from free-form letters [53], are treated as an existing partial correction in Section 10 rather than as a case. And the extensive evidence on differential language in letters [56, 35, 48] bears on the distribution of outcomes rather than on the structure of the instrument, and is reserved for Section 11.

4.3 Conditions of Disconfirmation

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

First, if divergence among referees is shown to be error rather than standpoint, the reconstruction reading fails and only the diagnosis of collapse in Section 7.5 survives. This is the condition the paper is most exposed to, for the reasons set out in Section 3.2, and the available evidence leaves it open in either direction.

Second, if no quantity can be exhibited that is invariant across projections of one relation, then there is nothing for a reconstruction to recover, and the requirement of several references is better explained as redundancy against individual unreliability.

Third, if referees turn out not to differ systematically by relational position once individual idiosyncrasy is accounted for, then independence of position is not the property that matters and the first design criterion of Section 10 is false.

Fourth, if institutions that weight by authority make better decisions than those that do not, the argument of Section 7.4 is false, and the practice the paper diagnoses as collapse is instead a defensible heuristic.

Fifth, if collecting the candidate's own account is shown to degrade decisions, whether by introducing self-presentation that evaluators fail to discount or by displacing attention from other sources, the principal proposal of Section 10 should be withdrawn. Evidence that disclosure can worsen what it reports exists in an adjacent domain and is treated in Section 13.

5 The Generative Relational Account

This section states the account the remainder of the paper applies. It gives the framework in the compact form the argument needs, states how an object without a name is identified, applies that to what a relation generates, states the requirement of revisability, and derives the criteria an instrument of transfer must satisfy.

5.1 Co-Emergence of Subject, Meaning, and Value within Relations

The framework within which this paper works may be stated in one sentence: it is a theory of how subject, meaning, value, creation, and normativity co-emerge through generative relational processes. Three consequences of that statement are used here and no more of the framework is imported.

The first is that what a relation produces is not held by either party. A working relation generates undertakings, understandings, difficulties, and judgments that neither participant would have produced alone and that neither possesses afterwards in the way a person possesses a memory of a fact. This is what Section 2.3 named co-experience.

The second is that access to what a relation generated is positional. Each party stands in the relation from somewhere, and that standpoint fixes what of the relation is available to them. A supervisor is placed to observe what a supervisor is placed to observe. A collaborator, a junior colleague, and a counterparty are each placed differently, and the differences are not degrees of completeness along one dimension.

The third is that these are ontological claims about the object and not claims about the imperfection of witnesses. A party who reports only what their position afforded has not failed to report the relation. There is no position from which the whole of it is available, including the positions of the two parties themselves.

5.2 Reconstruction of Pre-Symbolic Objects from Their Presentations

An object may be identified before it is named, and the means of identifying it is what persists across the appearances from which it is inferred.

The clearest instances are formal. An attractor in a dynamical system is not any trajectory and is not the average of trajectories; it is identified by what remains invariant across them, and the formal result that a reconstruction from observations preserves the topological invariants of the original is the statement of that idea in its exact form [52]. An optimum in an allocation problem is likewise not any allocation, and is identified by a property that holds across the allocations that exhibit it.

The formal results carry conditions and the paper claims none of them. The reconstruction theorem requires a deterministic, smooth, finite-dimensional system, a generic observable, and a sufficient embedding dimension, and a relation between two people satisfies none of these. What transfers is the form of identification and not the guarantee: an object may be recognised through what its presentations have in common before any presentation is taken to be the object.

Two literatures supply the disciplined version of the same idea. Measurement models posit an unobserved variable standing behind several observed indicators, and the argument that such models require realism about the unobserved variable is the argument that the indicators are indicators *of* something rather than a summary *into* something [4]. And the philosophical account of robustness holds that what can be detected by several *independent* means is more trustworthy and more likely to be real than what rests on one, with the emphasis falling on independence, since a second determination that shares the failure modes of the first adds nothing [62].

Definition 5.1 (Presentation and invariant). A *presentation* of an object is an appearance of it available from one standpoint. An *invariant* is a property exhibited by presentations from standpoints that differ in the respects that would be expected to alter it.

Definition 5.1 makes the independence requirement part of the definition. A property common to two presentations from the same standpoint is not an invariant in this sense, since the standpoints did not differ in the respect that would test it.

5.3 Co-Experience and the Positions from Which It Appears

Applying the preceding two subsections gives the paper’s central object.

Definition 5.2 (Projection). A *projection* of a co-experience is an account of it given from one relational position. A projection is complete with respect to that position and partial with respect to the co-experience.

Three consequences follow and each does work later.

A projection is not a lossy copy. The distinction matters because the two diagnoses recommend different remedies. If an account were a degraded copy, the remedy would be a better copy: a longer letter, a more careful writer, a structured form. If it is a projection, no improvement in fidelity supplies what a single standpoint did not afford, and the remedy is a second standpoint.

Divergence between projections is therefore expected rather than anomalous. The finding that two referees agree about one candidate less than one referee agrees with themselves across two candidates [2] is, on this account, close to what should be predicted. Whether it should instead be read as error is the question left open in Section 3.2 and taken up in Section 6.

And the candidate holds a projection. The candidate stood in the relation, from a position no other party occupied, and the account available from that position is a presentation of the same

object. Nothing in the ontology distinguishes it in kind from the others, and Section 10.2 draws the consequence.

5.4 Revisability and the Historical Emergence of Interpretation

An interpretation of what a relation generated is made at a time, from the positions then available, on the presentations then to hand. The framework holds that such interpretations must remain open to revision, and the ground of that requirement is worth stating rather than assuming.

Interpretations of co-experience are made under three conditions that change. The presentations available change, as parties who were not consulted become available and as parties reconsider. The standpoints from which the relation can be viewed change, since a relation continues to have consequences after the interpretation is fixed. And what the interpretation is for changes, since a judgment made for one purpose is later relied on for others.

Claim 5.3 (Revisability). An interpretation of a co-experience that is given durable institutional force and no channel of revision forecloses the further interpretation of that co-experience. Where the interpretation was formed from a subset of the available positions, the foreclosure operates on an object that was never reconstructed.

Claim 5.3 is the paper's normative commitment and it is narrow. It does not hold that every judgment must be reopened, that no decision may be final, or that a candidate is entitled to a favourable interpretation. It holds that fixing an interpretation and removing the means of revising it are two acts, and that an instrument which performs both should be assessed as performing both.

5.5 Criteria the Account Imposes on an Instrument of Transfer

The preceding subsections yield five criteria. They are stated here so that Section 7 through Section 10 assess the instrument against a standard fixed in advance rather than against observations gathered afterwards.

1. **Plurality of standpoint.** An instrument transferring an account of a relation should draw on presentations from standpoints that differ in the respects that would alter what is seen.
2. **Preservation of divergence.** An instrument should carry forward what differs among presentations rather than resolving the difference before the recipient sees it.
3. **Inclusion of the co-producers.** Every party in whose relation the co-experience was generated holds a presentation of it, and an instrument that omits one omits a standpoint.
4. **Correction for the position of the source.** An instrument should carry what is known about how each source's standpoint and history bear on what that source reports, since a presentation cannot be interpreted without knowing from where it was taken.
5. **Availability of revision.** An instrument giving an interpretation durable force should supply a channel by which that interpretation can be reopened.

Two remarks about the criteria. They are stated as requirements on the instrument and not on any person, and none of them requires a referee to be more diligent or more honest. And they are stated in a form that permits failure to be identified: Section 7 finds the instrument failing the second and fourth, Section 10 treats the first and third, and Claim 5.3 governs the fifth.

6 Reconstruction of Co-Experience through Multiple Reference

This section states the paper’s central reading and argues for it. The requirement of several references is read as an attempt to reconstruct a co-experience from its presentations. The argument that the divergence among those presentations carries information about the relation, rather than only about the referees, is made here rather than assumed, since Section 3.2 recorded that the field is divided.

6.1 Projection of Co-Experience from a Single Relational Position

A single reference is a projection in the sense of Definition 5.2: an account of a co-experience given from one relational position, complete with respect to that position and partial with respect to the relation.

Three features of the practice follow directly and are worth recording, because an account should first make the ordinary features of its object intelligible.

Institutions ask referees to state the capacity in which they knew the candidate, and treat that statement as material rather than as preliminary. On the reading offered here it is material: it identifies the standpoint from which the presentation was taken, without which the presentation cannot be placed.

Institutions prefer referees who knew the candidate differently. A programme that asks for a research supervisor, a teaching supervisor, and a collaborator is specifying standpoints rather than collecting opinions, and the preference is difficult to explain if what is wanted is the best-informed judgment, since in that case the best-informed judge should simply be asked twice.

And referees are chosen by the candidate. This is ordinarily treated as a weakness of the instrument, and under the reading offered here it is at least ambiguous, since the candidate is the only party who knows which standpoints exist. What the practice lacks is not the candidate’s knowledge of the set of positions but any requirement that the chosen positions differ.

6.2 Independence of Position and the Requirement of Multiple Reference

The requirement of several references is, on this reading, an instance of multiple determination, and the property that makes multiple determination worth anything is independence [62]. A second determination sharing the failure modes of the first adds confidence without adding information.

Claim 6.1 (Independence of position). What a reconstruction requires of its sources is that their standpoints differ in the respects that would alter what each is placed to observe. The number of accounts is a proxy for this and a poor one, since accounts from a common standpoint are one presentation repeated.

Claim 6.1 follows from Definition 5.1, which builds the requirement into what an invariant is, and it has an immediate practical consequence stated in Section 10.1: three letters from one laboratory satisfy a numerical requirement and fail a reconstructive one.

The multi-source literature supports the claim in the limited form that its evidence permits. Correlations between ratings from different sources are low while reliabilities within a source are markedly higher [9], which is what one expects if the source is doing work; and peer and subordinate ratings add validity beyond supervisor ratings [10], which is to say that they carry what the supervisor’s rating does not. Neither result establishes that the additional variance is standpoint. Both establish that additional sources carry what the first lacks, which is the weaker claim the design criterion needs.

6.3 Invariance across Projections

The paper’s central premise is that what identifies the object is what persists across the projections, and this subsection argues for it against the reading on which the differences are error.

Restatement of the two readings. Under the error reading, several referees are parallel measurements of one quantity, each contaminated by halo and by individual idiosyncrasy; the appropriate treatment is aggregation, which recovers the true score, and the divergence is what aggregation removes [59]. Under the standpoint reading, several referees are presentations of one object taken from positions that afford different things; the appropriate treatment is comparison, and the divergence is where the information lies.

Limits of what the numeric evidence settles. The decomposition of multisource ratings assigns most variance to the individual rater rather than to the person rated [50], and the strongest reanalysis in the standpoint tradition recovers a systematic source component of about eight per cent [30]. Both readings accommodate these numbers, because a component attached to the individual rater is exactly what a standpoint account predicts where standpoints are individually occupied, and is exactly what an error account predicts where raters are individually unreliable. The paper therefore rests its argument on other grounds.

Considerations bearing on the choice between the readings. Three considerations bear on it. (1) **The material differs.** The evidence on both sides comes from numeric ratings of a common set of performance dimensions, in which every rater is asked the same question about the same construct. A reference is a narrative account of a relation, in which the referee reports what the relation contained from where they stood, and the questions two referees answer are not the same question. Aggregation is the correct treatment of parallel measurements, and its correctness for accounts that were never parallel remains open.

(2) **The error reading requires a quantity of which the referees are measurements.** Halo is defined as the contamination of ratings of distinct dimensions by an overall impression [59], and the correction for it presupposes that the dimensions are dimensions of one thing possessed by the rated person. Where the object is what a relation generated, the presupposition is what the account of Section 5.1 denies: there is no single quantity possessed by the candidate of which a supervisor’s account and a collaborator’s account are two readings.

(3) **The two readings differ in what they predict about the structure of divergence, and the difference is testable.** If the divergence is idiosyncrasy, it should not be organised by relational position, and referees who stood in similar positions should differ from one another as much as referees who stood in different positions. If the divergence is standpoint, it should be organised: referees in similar positions should converge relative to referees in different positions, after individual idiosyncrasy is accounted for. The reanalysis that recovers a source component [30] is evidence of this organisation, and its magnitude is small. This paper leaves the test on reference material to further work, and Section 13 records that the central premise is argued and not demonstrated.

Consequences for the paper if the error reading is right. The paper is written so that this can be stated. If divergence among referees is idiosyncrasy, then the reconstruction reading fails, and the requirement of several references is best explained as redundancy against unreliable individual measurement. Two of the paper’s results survive that outcome. The diagnosis of collapse in Section 7.5 survives, since aggregating a set differs from reading the most authoritative member of

it, and under the error reading collapse is worse rather than better: it selects one noisy measurement instead of averaging several. And the argument of Section 7.4, that authority is anti-correlated with contact, survives unchanged. What fails is the treatment of divergence as signal and the design criterion that follows from it.

6.4 Relational Generativity and the Search for an Invariant

If the reconstruction reading is right, a question follows that the account has so far avoided: what is invariant across the projections?

The obvious answer is unavailable within the framework. Traits of the candidate cannot be the invariant, because the ontology of Section 5.1 holds that what a relation generates is generated in the relation, so that a property exhibited in one relation is not guaranteed to be exhibited in another. An account in which the referees triangulate onto a stable set of dispositions would be a measurement account wearing relational vocabulary.

The candidate answer this paper offers is different in kind and is stated as a conjecture rather than a result.

Claim 6.2 (Candidate invariant). What may persist across projections of a candidate's relations is not a set of properties the candidate carries but the candidate's characteristic effect on the generativity of the relations they enter: whether relations they join tend to become more or less able to produce further undertakings, claims, and revisions.

Claim 6.2 is proposed for three reasons. It is a property of relations rather than of persons, which is what the framework's ontology permits. It is a property that could in principle be exhibited across relations of different kinds, which is what an invariant must be. And it names something that referees in different positions might each have observed from their own position, which is what makes it recoverable from projections.

Three qualifications are recorded and none is answered here. No measure of the proposed quantity is offered. Whether anything is invariant across a person's relations remains open, and nothing may be, in which case the reconstruction has no object and the requirement of several references is doing something else. And if the claim is correct, it implies that current instruments ask referees the wrong question, since a referee asked whether a candidate is excellent is being asked about a property and not about an effect on a relation.

6.5 Generation of the Projection within a Second Relational System

A projection travels only when a party carries it. It is written, and the writing occurs within a second relation, that between the referee and the institution, which has its own standpoint, its own conventions, and its own stakes.

This is the point at which the literature on portable records bears. An inscription travels because it has been made mobile and immutable [33], and what survives the making is what the receiving context recognises. The pursuit of impersonal, transferable judgment arises where personal authority is weak [45], and the transformation of qualities into commensurable terms renders invisible whatever the common scale does not carry [20].

Claim 6.3 (Second-system generation). An account of a co-experience is generated within the relation between the referee and the receiving institution, and not within the relation it reports. What it carries is therefore selected by what the receiving relation recognises, and the selection is systematic rather than random.

Claim 6.3 explains a feature of the practice that the projection account alone leaves unexplained. Referees write in a register that inflates, so that almost every account is favourable [1]. That is not a property of what the referee observed and it is not idiosyncrasy either; it is a property of the second relation, in which an unfavourable account carries costs and a favourable one does not.

Two consequences are carried forward. The correction of Section 10 must act on the second relation as well as on the first, since improving what referees observe would leave the register untouched. And the calibration proposed in Section 8.5 is a correction of exactly this kind: it leaves the referee's writing untouched and records how that referee's accounts have stood up, which is information about the second relation that the recipient can use.

7 Mis-Specification of the Instrument

Section 6 read the requirement of several references as an attempt at reconstruction. This section identifies where the instrument departs from what that attempt requires. The departures are assessed against the criteria fixed in Section 5.5, and the section finds the instrument failing the second and the fourth.

7.1 Endorsement, Reputational Collateral, and Contingent Liability

A referee who writes favourably does two things at once, and separating them is necessary before the rest of the section can proceed.

The referee supplies an account of what they observed. The referee also stakes something: a named person, identifiable and locatable, has associated themselves with a prediction about a candidate, and if the candidate performs badly the association is available to be recalled. The second act stands apart from the first. An anonymous account of identical content would supply the same observation and stake nothing.

The staking is what makes the account usable by a party who cannot verify it. A selection committee that lacks any means of checking what a referee reports is not in a position to weigh the observation, and is in a position to weigh what the referee has put at risk in reporting it. The committee is thereby relieved of a burden: a decision taken on the strength of a named person's endorsement is defensible in a way that a decision taken on the committee's own reading of the evidence is not.

Neither of these observations is a criticism, and the section states them because they explain what follows. An instrument that carries both an observation and a stake will be read for whichever of the two the reader can use, and a reader who cannot verify observations will read it for the stake.

7.2 Epistemic Weight and Decision Weight

Two quantities are therefore in play and they need not coincide.

The *epistemic weight* of an account is what it contributes to knowing what the relation contained: a function of how much of the relation the referee's position afforded, over what period, under what conditions, and how much of that the account reports.

The *decision weight* of an account is what it contributes to the decision actually taken: a function of what the reader can do with it, which includes what the reader can verify, what the reader can defend, and what the reader can attribute.

The instrument carries no marking that distinguishes them. A letter states neither how much of the relation its writer was placed to see nor what the writer is staking, and a reader must infer

both from the same cue, namely who the writer is. The two quantities are then read off one signal, and the signal is the one that tracks decision weight.

This is a failure of the fourth criterion of Section 5.5. An instrument should carry what is known about the position from which each account was taken, because a presentation cannot be interpreted without knowing from where. The instrument carries the identity of the source, which is a proxy for standing and not for position.

7.3 Separate Relational Origins of Authority and of Contact

Standing and position differ in origin as well as in kind: they are generated in different relations, which is why one is a poor proxy for the other.

A referee's standing is generated in that referee's relations with a field: it accumulates through publication, appointment, the holding of office, and the recognition of peers. A referee's contact with a candidate is generated in the relation between them: it accumulates through joint work, supervision, shared difficulty, and time.

Nothing connects the two. A person may accumulate a great deal of the first and very little of the second with respect to any given candidate, and the conditions that produce the first tend to reduce the second, since the roles that confer standing consume the time in which contact would be made.

Section 6.5 named the general form of this: an account is generated in the relation between referee and institution and carries what that relation recognises. Standing is what that second relation recognises most readily, because it requires no verification. Position is what the first relation contains, and the instrument transmits it only if the referee chooses to state it.

7.4 Thinness of High-Authority Projections

The preceding subsection has a consequence for weighting that runs against the practice.

Claim 7.1 (Thinness). Symbolic authority accrues to those whose time is most contested, and contact with a particular candidate consumes time. A referee of high standing therefore tends to supply a projection taken from a more distant position than a referee of low standing, and to that extent a thinner one.

Claim 7.1 concerns a tendency and not a rule. A senior researcher who has worked closely with a candidate for years supplies both standing and position, and the claim allows that case. It holds that the two vary independently, and that where they trade off, the trade runs in this direction.

The literature on expert judgment supplies a parallel finding at the level of accuracy rather than access. In a programme collecting some twenty-eight thousand predictions from two hundred and eighty-four experts over nearly two decades, forecasters were often only slightly more accurate than chance and were beaten by simple extrapolation, and the experts who were most visible and most confident were among the worst calibrated [54]. The finding is about a different quantity from the one this section concerns, and it is consistent with the direction of the argument: eminence is not a guide to accuracy, and it may be a countersignal.

The consequence for a reconstruction is stronger than the consequence for a prediction. A set of projections weighted by the standing of their authors is a set weighted toward the most distant standpoints, which is the least suitable weighting available for recovering an object from its presentations. On the reading of Section 6, weighting by authority is not a rough heuristic that could be improved; it is a weighting that works against the operation the instrument is performing.

7.5 Collapse onto the Most Authoritative Projection

The failure this section identifies as central lies in what is done with the material rather than in what is collected.

An institution requires three accounts, obtains them, and decides on the strength of the one signed by the most eminent name. The remaining two are read, and they do not bear on the outcome unless they contain something disqualifying. The reconstruction is thereby abandoned at the point of use.

Three things follow, and they compound.

The institution has paid the cost of reconstruction and taken none of its benefit. It has imposed on the candidate the work of securing several referees, and on the referees the work of writing, and has then used one account.

What it has collapsed onto is, by Claim 7.1, the projection taken from the most distant position. The collapse is therefore not a random selection from the set but a selection biased toward its least informative member.

And the collapse is invisible in the record. An institution that collected three letters and used one has a file indistinguishable from an institution that collected three and reconstructed from them, since nothing in the procedure records how the accounts were weighed. This is why the failure persists without correction, and why the remedy proposed in Section 10.4 is a requirement to record rather than an exhortation to weigh differently.

Collapse is a failure of the second criterion of Section 5.5, which requires that an instrument carry forward what differs among presentations rather than resolving the difference before the recipient sees it. Collapse resolves it maximally, by discarding all but one.

The diagnosis does not depend on the contested premise of Section 6.3, and this is worth stating since it is the paper's most robust result. If divergence among referees carries information, collapse discards it. If divergence is idiosyncratic error, the correct treatment is aggregation, and collapse is a failure of that too, since selecting one measurement from several is precisely what aggregation exists to avoid. Under either reading, deciding on the most authoritative single account is the worst available use of a set of accounts.

8 Issuance, Debasement, and the Absence of Calibration

Section 7 treated the accounts. This section treats the position of the party who issues them, and identifies the omission that Section 5.5 named as the fourth criterion: an instrument should carry what is known about how each source's standpoint and history bear on what that source reports.

8.1 The Referee's Position in the Instrument

The referee occupies three positions at once and the instrument distinguishes none of them.

The referee is a *witness*, having stood in the relation and observed what that position afforded. This is the position the reconstruction account cares about, and it is the one about which the instrument records least.

The referee is a *guarantor*, staking a reputation that the receiving institution can locate and recall, as Section 7.1 set out. This is the position the receiving institution can use without verification, and it is the one the instrument records most legibly, since the signature carries it.

And the referee is a *party to a continuing relation* with the candidate. An account is written by someone who will meet its subject again, whose own standing is affected by the candidate's later

performance, and who may have obligations toward the candidate arising from the relation itself. This position is recorded nowhere and it bears on everything the referee writes.

The three positions are not in conflict in any particular case. They are simply distinct, and an instrument that renders them by a single signature supplies no means of telling which is doing the work in a given account.

8.2 Historical Accumulation of a Signature's Worth

The weight a signature carries is not a property of the person. It is accumulated within a field over time, through the same processes that produce standing generally, and it is held by the field rather than by the signatory. The consequence is that a signature's worth can rise and fall without any change in what its holder observes or reports.

Two implications follow for the paper's argument.

The first is that the weight attaching to an account is historical in the sense Section 5.4 used: it is the residue of a history of recognition, and it is revisable in principle by the same field that conferred it. Nothing about it is fixed by the relation the account reports.

The second is that a field which confers weight through recognition alone is conferring it on a basis unconnected to accuracy. The finding that visibility and confidence run against calibration [54] is what one expects where the conferring process tracks something other than the property being credited.

8.3 Overissuance and the Debasement of a Signature

A referee who writes favourably for every candidate conveys nothing by writing favourably for one. The point is elementary and its consequences are not.

The corpus evidence indicates that this is the ordinary condition rather than an abuse: of nearly seven thousand reference ratings, ninety-six per cent rated candidates above average and fewer than one per cent below [1]. The signal is issued at nearly full strength for nearly everyone, which is the definition of a debased signal.

Claim 6.3 explains why this is stable rather than self-correcting. The register is a property of the relation between referee and institution, in which an unfavourable account carries costs to the referee that a favourable one does not: it damages a continuing relation with the candidate, it invites dispute, and it may be attributed. A referee who wrote at variable strength would bear those costs individually while the benefit, a more informative signal, would accrue to a system of which the referee is one participant among thousands.

Two features of the resulting equilibrium bear on the corrections proposed later. No individual referee can escape it by writing differently, since a single unusually candid account is read as a condemnation rather than as calibration. And the inflation falls unevenly, since what counts as the ordinary register differs across fields, countries, and languages, which Section 11 takes up as a distributional matter rather than a technical one.

8.4 Returns to Issuance and the Obligation of the Endorsed

Issuing an endorsement carries a cost to the issuer and yields a return to them, and the instrument records neither.

The costs are the ones just described: a stake placed at risk, and the effort of writing. The returns are less often stated. A referee whose endorsements are accepted acquires evidence that their endorsements are accepted, which is itself a component of standing. A referee accumulates, through the practice, a set of former candidates who have reason to regard the referee favourably.

And a referee occupying a position through which candidates must pass acquires influence over their trajectories that no office confers.

This paper records these returns without developing them. The exchange between a referee and a candidate, in which an account is given and an obligation incurred, and the accumulation of standing through the issuing of accounts, are the subject of the political-economy paper in this series and are not argued here. What the present section needs is the narrower point that the referee has a position in the instrument that is neither witness nor guarantor, and that the instrument records it no better than it records the others.

8.5 Calibration of Projections and the Absent Clearing House

The omission this section has been building toward can now be stated.

A projection can be interpreted only if something is known about the standpoint from which it was taken and about the systematic distortion that standpoint introduces. Nothing in the practice supplies this. No institution maintains, and no institution has access to, a record of how a given referee's past accounts have stood up: whether the candidates that referee described as exceptional performed exceptionally, whether that referee's accounts discriminate among candidates at all, or how that referee's register compares with the register of others in the same field.

Claim 8.1 (Absent calibration). The instrument transmits accounts without any record of the accuracy or the discrimination of their sources. A recipient therefore cannot correct for the distortion a given source introduces, and cannot distinguish a source whose favourable account is informative from one whose accounts are uniformly favourable.

The apparatus for correcting this exists and is old. Forecast accuracy has been scored by a proper scoring rule since the middle of the last century [7], and the demonstration that individual forecasting accuracy is stable enough across time to be a property of the forecaster rather than of the occasion is established in the tournament literature [55, 54]. The evidence that eminence and calibration diverge [54] is precisely the evidence that calibration would have to be measured rather than inferred from standing.

Three features of the proposal are worth stating here and are taken up in Section 10.

Calibration is a correction applied to the source and not a demand made of the source. It asks referees to write nothing differently. It gives the recipient what is needed to read what referees write.

Calibration acts on the second relation identified in Claim 6.3. It leaves untouched the register in which referees write, which no individual referee can change, and records that register, which makes it interpretable.

And calibration is the only one of the paper's proposals that would make the debasement of Section 8.3 self-correcting, since a referee whose accounts ceased to discriminate would be recorded as not discriminating, and the cost of uniform favourability would fall on the party producing it rather than on the candidates it fails to distinguish.

One further finding bears directly on whether the proposal would work. Reputational discipline of an assessor has been examined formally, with the conclusion that the reputation argument holds only where a sufficiently large fraction of the assessor's income comes from sources other than the assessments in question [36]. Transposed, a referee whose standing depends substantially on the placements their accounts secure is not disciplined by having those accounts recorded, and the discipline works where the referee's position rests on other things. The condition is stateable and holds unevenly.

Section 13 records the hazard the proposal creates, which is that a record of referee accuracy is itself an instrument of power over referees, and Section 9.3 examines a case in which the party assessing was itself assessed and the assessment did not produce accuracy.

9 Epistemic Compression and the Delegation of Judgment

The preceding sections treated the accounts and their sources. This section treats the receiving institution, and identifies a consequence of relying on references that operates over time rather than in any single decision.

9.1 The Compression Cascade in Selection

An institution selecting among many candidates has more assessments to make than it has capacity to make them. References supply a way of proceeding: a judgment is available from someone who had access the institution lacks, and accepting it is cheaper than generating one.

The acceptance has a consequence for the next round. An institution that decides on referees' judgments does not, in that round, exercise its own judgment about persons, and therefore does not develop it. Its assessments in the following round are made with no more capacity than before and with the same shortage of time, so the same expedient recommends itself more strongly. The reliance is self-reinforcing, and the direction of travel is toward deciding on the strength of judgments the institution is progressively less able to evaluate.

Nothing in this is a failure of rationality by the institution. Relying on another party's judgment where that party had access one lacks is the ordinary condition of divided cognitive labour, and it is rational rather than defective [29]. The difficulty lies in what sustained reliance does to the capacity that reliance was supposed to supplement.

9.2 Degradation of an Institution's Capacity to Generate Knowledge of Persons

The structure of that difficulty is documented in another setting.

Automating the routine portions of a task removes the occasions on which the operator would have exercised the corresponding skill, with the result that the operator is least able to intervene at the moment when intervention is required; the irony is that the automation which removed the easy parts of the work leaves the operator responsible for the hardest part and least practised at it [3]. Complacency and bias in the use of automated aids have a common attentional basis, appear in experts as well as in novices, and are not removed by training or by awareness of the effect [44]. The general industrial form of the same structure is the separation of conception from execution and the consequent loss of the capacity to conceive [6].

Claim 9.1 (Capacity degradation). An institution that decides on the judgments of referees loses, over time, not information about candidates but the capacity to generate knowledge of persons of its own. The loss is of a capacity rather than of a stock, and it is therefore not remedied by obtaining more accounts.

Claim 9.1 follows from the account of Section 5.1 rather than from the automation analogy alone, and the derivation is worth stating. Knowledge of what a person generates in relations is itself generated in a relation. An institution acquires it by standing in some relation to the person: through work performed together, through extended assessment, through observation over time. An institution that never enters such a relation has no position from which anything could be generated,

and its knowledge of persons consists entirely of accounts of relations it was not party to. It is not merely uninformed; it is not placed.

Two consequences are recorded. The remedy lies elsewhere than in more references, since further accounts of relations the institution stood outside leave its position unchanged. And an institution in this condition is least able to perform the correction proposed in Section 8.5, since evaluating whether a referee's past accounts stood up requires knowing how the candidates described in them subsequently fared, which requires the institution to have been in a position to know.

9.3 Credit Rating Agencies

The largest documented instance of this structure occurred in financial markets and is instructive both for the cascade and for what happened when it was reformed.

Delegation of assessment to the rating agencies. Assessments of creditworthiness were delegated to a small number of rating agencies, and the delegation was entrenched by regulation. The Securities and Exchange Commission declared that only the ratings of nationally recognised statistical rating organisations were valid for determining broker-dealer capital requirements; other financial regulators adopted the same category; and regulators thereby delegated their safety decisions to the rating agencies [60]. The regulatory structure propelled three agencies to the centre of the bond markets and, in doing so, virtually guaranteed that when those agencies made mistakes the mistakes would have serious consequences for the financial sector [60]. An overview of the industry's structure and regulation is available in the same author's later survey [61].

The structural resemblance to the selection case is close. An assessment that each institution could in principle have made for itself was made instead by an external party; reliance was reinforced by the convenience and the legitimacy of the external assessment; and the capacity to assess independently was not maintained.

The reform and its measured effect. The legislative response addressed the reliance directly. Section 939A of the Dodd-Frank Act required every federal agency to review its regulations requiring an assessment of creditworthiness, to remove references to and requirements of reliance on credit ratings, and to substitute a standard of creditworthiness the agency deemed appropriate [16]. The remedy was therefore not to improve the assessor but to stop making the assessment load-bearing.

What followed is the finding this paper most needs. An empirical assessment of the Act's effect on ratings found no evidence that it disciplined the agencies toward accuracy; ratings became lower, false warnings more frequent, and downgrades less informative, consistent with agencies protecting their own reputations rather than improving their assessments [15].

Lessons carried forward from the rating-agency case. (1) **Removing the load-bearing status of an external assessment is a coherent intervention, and it was the one the legislature chose.** The parallel in Section 10 is to place references in a diagnostic role rather than a ranking one.

(2) **Subjecting an assessor to scrutiny is a caution against the paper's own proposal.** Subjecting an assessor to scrutiny changed the assessor's behaviour in a direction that protected the assessor rather than improving the assessment [15]. The calibration proposed in Section 8.5 is a form of scrutiny of assessors, and this case indicates that such scrutiny may produce defensive rather than accurate accounts. Section 13 records this as the strongest available objection to that proposal.

9.4 The Epistemic Duty of Selection

The preceding subsections support a normative claim, and it is stated carefully.

Claim 9.2 (Epistemic duty of selection). An institution whose decision materially affects a person's trajectory owes an epistemic effort proportionate to that effect, and the effort is owed to knowing the person rather than to obtaining defensible grounds for a decision about them. Substituting the second for the first satisfies the institution's requirements and not the duty.

Claim 9.2 takes a familiar form. Institutions accept scaled epistemic obligations routinely where the consequence is financial: an acquisition, an investment, or an extension of credit attracts an inquiry proportioned to the sum at stake, and no one argues that resources make the inquiry optional. What the claim adds is the observation that the same institutions accept no comparable obligation where the consequence is a person's trajectory, and that the asymmetry has no evident justification.

The claim is bounded in three ways. It concerns effort and not accuracy, since an institution that inquires proportionately may still decide wrongly. It is proportionate rather than absolute, so a decision with small consequence attracts small effort. And it does not require the institution to enter a relation with every candidate, which is impossible; it requires that the institution's inquiry not consist entirely of accounts of relations it was not party to, which the sequential arrangement proposed in Section 10.5 makes feasible at the point where consequence is concentrated.

10 Correction of the Instrument

This section states what would follow from taking the reconstruction reading seriously. The proposals are assessed against the criteria of Section 5.5 and against the evidence that measures of this kind sometimes fail. None of them requires a referee to write better.

10.1 Independence of Relational Position among Sources

The first correction follows from Claim 6.1 and satisfies the first criterion.

Institutions should specify the *positions* from which accounts are wanted, and should treat a set of accounts from a common position as one account. A requirement of three letters is satisfied by three colleagues from one laboratory who observed the candidate in the same capacity, under the same conditions, in the same period. On the reconstruction reading, that set contains one presentation recorded three times, and its apparent agreement is not evidence of anything, since the standpoints did not differ in the respects that would have tested the agreement.

Two implementations are available and they differ in cost. The weaker asks the referee to state the position occupied, the period, the conditions, and the respects in which the referee's view of the candidate's work was partial. The stronger specifies the positions in advance, as some institutions already do when they ask for a research supervisor, a teaching supervisor, and a collaborator, and treats an unfilled position as an unfilled requirement rather than as a preference.

The limit of this correction is that it depends on positions being available. Section 11.1 treats candidates for whom they are not.

10.2 Required Collection of the Candidate's Own Projection

The second correction follows from the third criterion and is the paper's principal proposal.

The candidate stood in each relation from a position no other party occupied, and holds a presentation of the same object. That presentation is currently collected from no one. The proposal is that it be collected as a required source.

Four features of the proposal are constitutive rather than incidental.

Structured and scored form of the account. (1) **The account is structured and scored, not narrative.** An unstructured self-narrative predicts weakly and adds little over other measures [42]. A structured self-report in which the subject describes actual accomplishments against defined dimensions, and in which those descriptions are scored by trained raters, predicts professional performance and is largely uncorrelated with cognitive ability [31]. The proposal is an instrument of the second kind. What is asked for is the position, the period, the conditions, the resources available, the responsibilities held, the difficulties encountered, and what the candidate takes the relation to have produced.

Collection before the third-party accounts arrive. (2) **The account is collected before the other accounts are received.** The request is made at the time referees are approached and closed before their accounts are received. This is what distinguishes the proposal from a right of reply, and it removes the objection that the candidate would tailor an account to what has been said, since nothing has yet been said. The sequencing is the same device as the one described in Section 10.3 and rests on the same reasoning [18].

Standing of the account among the other sources. (3) **The account is a source and not a defence.** The candidate's account enters the evidence base on the same footing as the others and is read against them. The purpose lies elsewhere than in allowing the candidate to answer a referee, and the referees' accounts remain closed to the candidate. What the candidate supplies is the conditions under which the relation took place, which is what the fourth criterion requires and what no other source is placed to give.

Expected agreement with the other sources. (4) **Low agreement with the other accounts is expected rather than disqualifying.** Self and supervisor ratings correlate at about .22 and self and peer at about .19 [9]. Under a measurement reading this is a reason to discount self-report. Under the reconstruction reading it is what projections from different positions should do, and the divergence is to be recorded rather than resolved, in accordance with Section 10.4.

10.3 Sequence Control and Protection against Collapse

The third correction addresses the failure identified in Section 7.5, and it is the proposal on which the evidence is most mixed.

The procedure is the one developed for expert examination under contextual bias: the primary material is analysed and the analysis recorded before the information that would colour it is supplied, with restrictions on revising the recorded analysis afterwards [18, 17]. Applied here, an evaluator reads the accounts and records an assessment before the identities of their authors are disclosed, and records separately any change made after disclosure.

The evidence requires two qualifications and the paper states them rather than claiming a settled remedy.

Removing an identity removes whatever that identity was carrying. A randomised comparison in peer review found that blinding lowered ratings and acceptance rates because it removed positive biases favouring authors from wealthy and English-speaking countries [26]; blinding therefore redistributes as well as neutralises. Whether blinding improves the quality of the resulting judgment is a further question on which trials disagree [57, 37]. And the best-known demonstration that

blinding changes outcomes carries acknowledged statistical qualifications, including estimates with large standard errors and one persistent effect in the opposite direction [28].

The proposal here is accordingly narrower than blinding and stops short of anonymisation. The identity of a referee is withheld only *until* the accounts have been read and an assessment recorded. What is being protected against is the collapse of a set of accounts onto its most authoritative member, and what is needed for that is that the reading of the accounts occur before the ranking of their authors is available, not that the ranking be unavailable.

The second qualification concerns what the identity legitimately carries. Section 7.2 distinguished the epistemic weight of an account from its decision weight, and the identity of the referee carries information relevant to the first, since a referee's position and history bear on how the account should be read. That information should be supplied *as such*, through the position statement of Section 10.1 and the calibration record of Section 8.5, rather than inferred from standing. The sequencing proposal and the calibration proposal are therefore complements: the first withholds the proxy until the accounts have been read, and the second supplies what the proxy was standing in for.

10.4 Recorded Findings of Divergence among Projections

The fourth correction follows from the second criterion and from the observation in Section 7.5 that collapse leaves no trace.

Institutions should be required to record, as part of the decision, where the accounts diverged and what the divergence was taken to indicate. The requirement is on the record and not on the conclusion: an evaluator remains free to find that the divergence indicates nothing, and is required to have noticed it and said so.

Three properties recommend the requirement.

It makes collapse visible. A file in which no divergence is recorded, from a set of accounts that diverged, shows that the set was not read as a set. This is what the present arrangement cannot show.

It is robust to the contested premise. Under the reconstruction reading, recorded divergence is where the information is. Under the error reading, recorded divergence is a measure of how unreliable the accounts were, and therefore of how much weight the set should carry. Both readings make the record useful, which is unusual among the paper's proposals.

And it costs little. It asks for a paragraph in a file that already exists, made at a moment when the evaluator has read the material.

10.5 Recommendation in a Diagnostic Rather than a Ranking Role

The fifth correction changes the question the instrument is asked to answer.

A reference used to rank candidates is asked which of them is better, which requires a common scale, and a common scale requires that the accounts be commensurable. Commensuration renders invisible whatever the scale does not carry [20], and on the reconstruction reading what it renders invisible is the divergence that identifies the object.

A reference used diagnostically is asked a different question: given that the candidate is under serious consideration, is there something about this relation that the institution should know. That question proceeds without a common scale, leaves the accounts incommensurable, and is answered by specific facts rather than by comparative judgments.

The rating-agency case supports the direction of the change. Section 9.3 recorded that the legislative response to excessive reliance on external assessment was to remove the assessment's

load-bearing status rather than to improve the assessor [16]. Placing references in a diagnostic role is the same operation.

Two consequences follow. Specific factual content becomes the object of the instrument, which is content a candidate could in principle contest, unlike a comparative judgment. And the register described in Section 8.3 matters less, since a uniformly favourable comparative judgment conveys nothing while a uniformly favourable answer to a specific factual question is informative if it is true.

10.6 Channels of Revision for Interpretations of Co-Experience

The sixth correction follows from Claim 5.3 and the fifth criterion, and it is the one this paper least succeeds in specifying.

An interpretation of a co-experience is fixed by a letter, given durable institutional force, and provided with no route by which it might be reopened. Under Claim 5.3 that combination forecloses the further interpretation of what the relation generated. Three partial measures follow from the paper's own materials and none is a channel of revision proper.

Recorded divergence, from Section 10.4, preserves within the record the material from which a further interpretation could be made, rather than resolving it at the point of decision.

The candidate's own projection, from Section 10.2, ensures that the record contains a presentation from the position of the party whose interpretation is otherwise absent.

And calibration, from Section 8.5, makes the sources themselves revisable, since a referee's weight becomes a function of a record that can change rather than of a standing that does not.

What none of these supplies is a means by which an interpretation, once acted on, can be reopened by the person it concerns. That requires standing, a forum, and a procedure, none of which the instrument contains and none of which this paper designs. Section 11.2 states the residue and Section 13 records the design as owed rather than delivered.

11 Injustice Surviving Correction

Section 10 proposed six corrections. This section states what remains wrong when all of them have been made. The section exists because a constructive paper that ends with its own proposals overstates them, and because three of the four items below are not remediable by any procedure this paper can specify.

11.1 Unequal Access to Positions of Observation

A reconstruction is only as good as the set of positions available to it, and positions are unequally distributed.

A candidate who has worked in a well-staffed laboratory, on collaborative projects, under several supervisors, with junior colleagues and external partners, can supply presentations from many standpoints. A candidate who worked alone, or in one relation, or under a single supervisor who has since become unavailable, cannot. The second candidate's reconstruction is thinner, and it is thinner for reasons that have nothing to do with what their relations generated.

Every correction in Section 10 makes this worse before it makes it better. Requiring independence of position (Section 10.1) imposes a requirement that the relationally poor candidate cannot satisfy. Recording divergence (Section 10.4) makes a thin set legible as thin. An instrument that reconstructs well from a rich set of positions and reports honestly that it cannot reconstruct from a poor one has improved its accuracy and worsened its distribution.

The adjacent finding in the literature concerns what selection rewards rather than what it can see. In elite professional hiring, employers sought candidates who were culturally similar to themselves in leisure pursuits, experiences, and self-presentation, and concerns about shared culture often outweighed concerns about absolute productivity [46]. The mechanism differs from the one this section describes, and the direction is the same: what a candidate brings to a selection is shaped by where the candidate has been.

This paper offers no correction. The observation belongs to the paper on unequal generative conditions in this series, where it is the subject rather than the residue.

11.2 Absence of the Co-Producer from the Reconstruction

Section 10.2 proposed that the candidate's account be collected as a required source, and that correction is real and limited.

What it supplies is a presentation from a position otherwise absent from the evidence. What it withholds is any authority over the interpretation built from the evidence. The candidate contributes a projection and does not participate in the reconstruction, does not see what the other sources said, does not learn what divergence was recorded, and holds no standing to contest what was concluded.

The co-experience was generated in a relation of which the candidate was one of two parties. The interpretation of it is made by a third party from accounts supplied by others, and the correction proposed here changes the evidence base without changing who interprets. Under Claim 5.3 that is a foreclosure, and Section 10.6 conceded that the paper supplies no channel of revision proper.

The legal position in one jurisdiction states the residue precisely. Federal regulation permits a student to waive the right to inspect confidential recommendations submitted on their behalf [21]. The waiver is available to the candidate, which is to say that the arrangement is formally consensual; what the candidate is offered is the choice between an account they may read and an account that will be believed. This paper leaves aside how that choice is presented in practice, and notes only that the provision makes non-inspection the ordinary condition rather than an incidental one.

Who holds authority to interpret a shared experience is the subject of the jurisprudence paper in this series. What the present paper establishes is that correcting the evidence base leaves it unanswered.

11.3 Co-Experience Inaccessible to Any Institutional Position

Some of what a relation generated is available from no position an institution can occupy or solicit.

Three kinds of case may be distinguished. There is what only the two parties know and neither will report, because reporting it would expose one of them. There is what neither party can articulate, and which shows only in what they subsequently do. And there is what is visible only to parties an institution will not approach, including those with whom the candidate had difficulty, and whose accounts are unobtainable precisely because the relation ended badly.

The consequence is stated as a claim because it bears on the paper's own proposals.

Claim 11.1 (Confidence and invisibility). An instrument that improves at reconstructing what is accessible to institutional positions becomes more confident about the object without becoming better informed about what those positions do not afford. Improved reconstruction therefore increases the risk of treating the reconstructed portion as the whole.

Claim 11.1 is a caution against this paper. The corrections of Section 10 would produce a better-founded assessment and a more warranted confidence in it, and nothing in them tells an institution

how much of the relation remained outside every position it consulted. The commensuration literature makes the general form of the point: what a common treatment does not carry becomes invisible rather than acknowledged as missing [20], and ordinal sorting presents the result as merit [24].

11.4 The Contestation Paradox and the Cost of Rupture

The final item is a paradox rather than a gap, and the paper states it without resolving it.

Suppose the channel of revision that Section 10.6 declined to design were built, so that a candidate could contest a referee's account before a body with standing to revise it. Exercising that right would be visible to the referee. The relation between them, which is the relation that generated the co-experience and that supplies the candidate's standing to speak about it, would be placed at risk by the act of speaking about it.

Adjacent evidence supports the shape of the exposure without settling its magnitude here. A survey of one thousand one hundred and sixty-seven employees examined what followed when those subject to interpersonal mistreatment voiced resistance, and found both work retaliation and social retaliation, with different forms of voice triggering different forms of retaliation depending on the social positions of the parties [11]. Two features transfer. The dependence on relative position is the asymmetry this subsection describes. And the same study reports health-related costs attaching to *silence*, that is, to enduring mistreatment without voicing resistance, which indicates that the alternative to contesting carries costs of its own. The setting differs from the one at issue here, and the finding is reported as adjacent rather than as evidence about candidates and referees.

The asymmetry is severe. The referee's position stands independently of the relation; the candidate's rests on it. A referee whose account is contested loses little. A candidate who contests may lose the relation, the future accounts it would have produced, and the standing within a field that the relation conferred. The right is therefore formally available and practically available only to candidates who can afford to lose the relation, which is a distributional property and not a procedural one.

Three observations bound the paradox without dissolving it.

The framework's requirement is openness rather than continuation: what must be preserved is the capacity of relations to continue being generated, not any particular relation. A relation that ends counts as a failure of generativity for no such reason, and the unit at stake is the field rather than the dyad.

By the paper's own criterion, a relation sustained by the impossibility of questioning what it produced is already in the condition Claim 5.3 identifies. The choice is not between preserving a relation and rupturing it; it is between a relation in which the interpretation of what it generated remains open and one in which it does not, and the second looks undisturbed because nothing in it can be disturbed.

And routing contestation through an institution rather than through the relation reduces the exposure without removing it. A claim addressed to a body that determines the matter differs from an accusation made to the referee, though the referee will learn of it.

What survives all three is the cost itself. Someone bears it, the framework places the decision with the party who bears it, and that is where this paper leaves it. The design of a forum in which the cost is lower belongs to the jurisprudence paper.

12 Implications for the Generative Relational Framework

Three results return to the framework, and two of them are corrections.

Pre-symbolic objects are identified through invariance, and the identification has conditions. The framework holds that what a relation generates is not held by either party and is available only positionally. This paper has taken the further step of asking how such an object is identified at all, and has answered that it is identified by what persists across presentations taken from positions that differ in the respects that would alter them (Definition 5.1). The independence requirement belongs to what an invariant is, rather than to methodological good manners. A framework that treats relational objects as real owes an account of how they are recognised, and this is one.

The framework’s own commitment to revisability has a cost it has not priced. Claim 5.3 holds that fixing an interpretation and removing the means of revising it are two acts. Section 11.4 shows that supplying the means has a cost borne asymmetrically by the party the framework intends to protect. The framework requires revisability and leaves open who pays for exercising it, and the answer cannot be that the party who bears the cost should be more courageous.

Accounts of relations are generated in second relations, and the framework has not treated this. Claim 6.3 holds that an account of a co-experience is generated within the relation between the reporter and the recipient, and carries what that second relation recognises. The framework has treated interpretation as an act performed on a relation. This paper finds that it is also an act performed within another one, and that the properties of the second relation, including what it makes costly to say, shape what the first can be known to have contained. That is a general feature and not a property of recommendation.

13 Limits of the Account

13.1 Conditions of Falsification

Section 4.3 stated five conditions and their status is as follows.

The central premise, that divergence carries information about the object, is argued in Section 6.3 and not demonstrated. The paper concedes that the variance decompositions accommodate both readings, that the systematic source component recovered by the strongest supportive study is approximately eight per cent [30], and that the contrary reading is held by a substantial programme [59]. The test the paper proposes, whether divergence is organised by relational position after individual idiosyncrasy is accounted for, is left to further work.

No quantity has been exhibited that is invariant across projections of one relation. Claim 6.2 proposes a candidate and offers no measure of it, and Section 6.4 records that nothing may be invariant, in which case the reconstruction has no object.

Whether referees differ systematically by relational position is untested on reference material. The evidence relied on comes from numeric ratings of common performance dimensions, which Section 6.3 argues is different material.

Whether institutions that weight by authority decide worse is untested. The argument of Section 7.4 is structural, and the parallel evidence on expert calibration [54] concerns accuracy rather than access.

A further hazard attaches to calibration specifically. Where a public measure is applied to a party, that party alters conduct in response to being measured, through self-fulfilling prophecy and through commensuration, so that the measure changes the world it reports [19]. A record of referee accuracy is such a measure, and referees may be expected to write so as to score well on it rather

than so as to report what they observed. The paper proposes the record and states this as its principal unquantified risk.

And whether collecting the candidate’s account would degrade decisions is untested. The strongest adjacent evidence is unfavourable in an important way: subjecting an assessor to scrutiny in the credit-rating case produced defensive rather than accurate assessment [15], and the calibration proposal of Section 8.5 is a form of scrutiny of assessors. The proposal is advanced with that finding against it.

13.2 Claims Advanced Without Support

Three claims rest on argument alone.

The contestation paradox of Section 11.4 is developed from the structure of the relation and not from evidence about what happens to candidates who contest.

Claim 11.1, that improved reconstruction increases confidence without increasing coverage, is an inference from the account and awaits measurement.

And Claim 6.2 is offered explicitly as a conjecture.

13.3 Sources Not Yet Verified

This draft cites only sources verified against a publisher, journal, index, or institutional page before the section using them was written. Several literatures the argument touches are therefore represented thinly or not at all, and the absences are substantive.

The formal economics of certification is represented by one result, on the conditions under which reputational concern disciplines an assessor [36], and the wider literature on ratings shopping and on competition among certifiers is absent. Work on unequal access to well-placed referees through professional networks is absent from Section 11.1, which relies on a single adjacent study [46]. Empirical work on retaliation against those who complain against superiors is now present in adjacent form [11] and concerns workplace mistreatment rather than contested accounts; no study of what follows a candidate’s contesting a referee’s account was located. Work on reactivity, in which those measured alter their conduct to satisfy the measure, is absent, and it bears on whether calibration would change what referees write. And research on the practice surrounding confidentiality waivers is absent, which is why Section 11.2 states the legal provision and declines to characterise the practice. Work on reactivity is now present [19] and is applied against the paper’s own calibration proposal in Section 13.1.

13.4 Extensions

Four extensions are identified and none is attempted.

An empirical test of whether divergence among referees is organised by relational position would settle the paper’s central premise in one direction or the other, and is feasible on existing reference corpora.

A measure of the quantity proposed in Claim 6.2 would convert a conjecture into a testable proposition, and would also indicate what referees should be asked.

A trial of the sequencing procedure of Section 10.3 in a selection setting would establish whether reading before authorship changes outcomes, and the peer-review literature indicates that it would [26] without indicating whether the change is an improvement.

And a design for a forum in which an interpretation can be reopened at a cost the affected party can bear is what Section 10.6 and Section 11.4 jointly require and what this paper does not supply.

14 Conclusion

Institutions that select by reference ask for more than one, and the requirement is difficult to explain if what is wanted is an opinion. This paper has read it as an attempt to reconstruct an object from its presentations. What a relation generates is held by neither party and is available only from the positions the parties occupied, so an account of it is a projection rather than a copy, and several projections are what a reconstruction requires. On this reading the institution is attempting something structurally correct.

It is executing it badly, and the failures are specific. Independence of position, which is what a reconstruction needs, is proxied by the number of letters, which is a different quantity. Divergence among accounts, which is where the information would lie, is resolved before the decision-maker sees it. The identity of the referee is made to carry both the epistemic weight of an account and its usefulness in defending a decision, and those quantities are generated in different relations and are not correlated. Authority accrues to those whose time is scarce, so weighting by authority weights toward the most distant standpoints. And the decisive failure is the collapse of a set of accounts onto its most authoritative member, which discards the reconstruction, selects its least informative element, and leaves no trace in the record.

The corrections follow from the diagnosis and none of them asks a referee to write better. Positions should be specified rather than letters counted. Divergence should be recorded rather than resolved. Referees should be calibrated, so that what a source's account is worth becomes a matter of record rather than of standing. The reading of accounts should precede the disclosure of their authors. References should be asked a diagnostic question rather than a ranking one. And the candidate, who stood in the relation from a position no one else occupied, should supply an account as a required source, structured and scored, collected before the other accounts arrive.

What survives the corrections is stated because a constructive paper that ends with its proposals overstates them. Positions of observation are unequally distributed, and an instrument that reconstructs honestly from a rich set and reports honestly that it cannot reconstruct from a poor one has improved its accuracy and worsened its distribution. The candidate contributes a projection and still holds no authority over the interpretation built from it. Some of what a relation generated is available from no position an institution can occupy, and an instrument that improves at recovering the rest becomes more confident without becoming better covered. And a candidate who contests an account risks the relation that generated the standing from which the contest is made, which is a cost that someone bears and that no procedure described here removes.

The account is a revisable proposal, and its central premise is contested in the literature it draws on. If divergence among referees is standpoint, the paper's reading holds and its corrections follow. If divergence is error, the reading fails, the diagnosis of collapse survives and is strengthened, and the requirement of several references is redundancy against unreliable measurement rather than an attempt at reconstruction. The paper is written so that a reader persuaded of the second position can identify precisely what to discard.

Acknowledgments

The present definitions, constructions, arguments, conclusions, and errors remain the author's responsibility. The interest arising from the author's own position with respect to the procedures examined here is declared in the front matter.

References

- [1] Michael G. Aamodt. Validity of recommendations and references. *Assessment Council News*, pages 4–6, February 2006. Reports Aamodt and Williams (2005).
- [2] Michael G. Aamodt, Devon A. Bryan, and Alan J. Whitcomb. Predicting performance with letters of recommendation. *Public Personnel Management*, 22(1), 1993.
- [3] Lisanne Bainbridge. Ironies of automation. *Automatica*, 19(6):775–779, 1983.
- [4] Denny Borsboom, Gideon J. Mellenbergh, and Jaap van Heerden. The theoretical status of latent variables. *Psychological Review*, 110(2):203–219, 2003.
- [5] Denny Borsboom, Gideon J. Mellenbergh, and Jaap van Heerden. The concept of validity. *Psychological Review*, 111:1061–1071, 2004.
- [6] Harry Braverman. *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*. Monthly Review Press, New York, 1974.
- [7] Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950.
- [8] Donald T. Campbell and Donald W. Fiske. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2):81–105, 1959.
- [9] James M. Conway and Allen I. Huffcutt. Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance*, 10(4):331–360, 1997.
- [10] James M. Conway, Karen Lombardo, and Kelly C. Sanders. A meta-analysis of incremental validity and nomological networks for subordinate and peer ratings. *Human Performance*, 14:267–303, 2001.
- [11] Lilia M. Cortina and Vicki J. Magley. Raising voice, risking retaliation: Events following interpersonal mistreatment in the workplace. *Journal of Occupational Health Psychology*, 8(4):247–265, 2003.
- [12] Lee J. Cronbach, Goldine C. Gleser, Harinder Nanda, and Nageswari Rajaratnam. *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. Wiley, New York, 1972.
- [13] Norman K. Denzin. *The Research Act: A Theoretical Introduction to Sociological Methods*. McGraw-Hill, 2 edition, 1978. First edition Aldine, 1970.
- [14] Paul J. DiMaggio and Walter W. Powell. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2):147–160, 1983.
- [15] Valentin Dimitrov, Darius Palia, and Leo Tang. Impact of the Dodd-Frank act on credit ratings. *Journal of Financial Economics*, 115(3):505–520, 2015.

- [16] Dodd-frank wall street reform and consumer protection act. Pub. L. No. 111-203, 124 Stat. 1376 (2010), §939A, 2010. Section 939A requires every federal agency to review regulations requiring an assessment of creditworthiness, remove references to and requirements of reliance on credit ratings, and substitute a standard of creditworthiness the agency deems appropriate.
- [17] Itiel E. Dror and Jeff Kukucka. Linear sequential unmasking–expanded (LSU-E): A general approach for improving decision making as well as minimizing noise and bias. *Forensic Science International: Synergy*, 3:100161, 2021.
- [18] Itiel E. Dror, William C. Thompson, Christian A. Meissner, Irving Kornfield, Dan Krane, Michael Saks, and Michael Risinger. Context management toolbox: A linear sequential unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. *Journal of Forensic Sciences*, 60(4):1111–1112, 2015.
- [19] Wendy Nelson Espeland and Michael Sauder. Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113(1):1–40, 2007.
- [20] Wendy Nelson Espeland and Mitchell L. Stevens. Commensuration as a social process. *Annual Review of Sociology*, 24:313–343, 1998.
- [21] Family educational rights and privacy act. 20 U.S.C. §1232g; 34 C.F.R. Part 99, §99.12. §99.12 permits a student to waive the right to inspect confidential recommendations submitted for admission, employment or honours.
- [22] Nigel G. Fielding and Jane L. Fielding. *Linking Data*, volume 4 of *Qualitative Research Methods*. Sage, 1986.
- [23] Uwe Flick. Triangulation revisited: Strategy of validation or alternative? *Journal for the Theory of Social Behaviour*, 22:175–197, 1992.
- [24] Marion Fourcade and Kieran Healy. Classification situations: Life-chances in the neoliberal era. *Accounting, Organizations and Society*, 38(8):559–572, 2013.
- [25] Marion Fourcade and Kieran Healy. *The Ordinal Society*. Harvard University Press, 2024.
- [26] Charles W. Fox, Jennifer Meyer, and Emilie Aimé. Double-blind peer review affects reviewer ratings and editor decisions at an ecology journal. *Functional Ecology*, 37(5):1144–1157, 2023.
- [27] Diego Gambetta. *Codes of the Underworld: How Criminals Communicate*. Princeton University Press, 2009. ISBN 9780691152479.
- [28] Claudia Goldin and Cecilia Rouse. Orchestrating impartiality: The impact of “blind” auditions on female musicians. *American Economic Review*, 90(4):715–741, 2000.
- [29] John Hardwig. Epistemic dependence. *The Journal of Philosophy*, 82(7):335–349, 1985.
- [30] Brian J. Hoffman, Charles E. Lance, Bethany Bynum, and William A. Gentry. Rater source effects are alive and well after all. *Personnel Psychology*, 63(1):119–151, 2010.
- [31] Leaetta M. Hough. Development and evaluation of the “accomplishment record” method of selecting and promoting professionals. *Journal of Applied Psychology*, 69(1):135–146, 1984.

- [32] Nathan R. Kuncel, Ross J. Kochevar, and Deniz S. Ones. A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International Journal of Selection and Assessment*, 22(1):101–107, 2014.
- [33] Bruno Latour. Visualisation and cognition: Thinking with eyes and hands. In Henrika Kuklick, editor, *Knowledge and Society: Studies in the Sociology of Culture Past and Present*, volume 6, pages 1–40. JAI Press, 1986. Reprinted, with revision, in Lynch and Woolgar (eds.), *Representation in Scientific Activity*, MIT Press, 1990, pp. 19–68.
- [34] Niklas Luhmann. *Trust and Power*. John Wiley, 1979. Expanded edition, Polity Press, 2017.
- [35] Juan M. Madera, Michelle R. Hebl, and Randi C. Martin. Gender and letters of recommendation for academia: Agentic and communal differences. *Journal of Applied Psychology*, 94(6):1591–1599, 2009.
- [36] Jérôme Mathis, James McAndrews, and Jean-Charles Rochet. Rating the raters: Are reputation concerns powerful enough to discipline rating agencies? *Journal of Monetary Economics*, 56(5):657–674, 2009.
- [37] Robert A. McNutt, Arthur T. Evans, Robert H. Fletcher, and Suzanne W. Fletcher. The effects of blinding on the quality of peer review: A randomized trial. *JAMA*, 1994. PMID 8015127.
- [38] John W. Meyer and Brian Rowan. Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2):340–363, 1977.
- [39] 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.
- [40] Jo Moran-Ellis, Victoria D. Alexander, Ann Cronin, Mary Dickinson, Jane Fielding, Judith Sleney, and Hilary Thomas. Triangulation and integration: Processes, claims and implications. *Qualitative Research*, 6(1):45–59, 2006.
- [41] Kevin R. Murphy. Explaining the weak relationship between job performance and ratings of job performance. *Industrial and Organizational Psychology*, 1:148–160, 2008.
- [42] Shane C. Murphy, David M. Klieger, Matthew J. Borneman, and Nathan R. Kuncel. The predictive power of personal statements in admissions: A meta-analysis and cautionary tale. *College and University*, 84(4):83–88, 2009. ERIC EJ851069; trade journal, no DOI.
- [43] National Research Council. *Strengthening Forensic Science in the United States: A Path Forward*. National Academies Press, Washington, DC, 2009.
- [44] Raja Parasuraman and Dietrich H. Manzey. Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3):381–410, 2010.
- [45] Theodore M. Porter. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press, 1995.
- [46] Lauren A. Rivera. Hiring as cultural matching: The case of elite professional service firms. *American Sociological Review*, 77(6):999–1022, 2012.

- [47] Paul R. Sackett, Charlene Zhang, Christopher M. Berry, and Filip Lievens. Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*, 107(11):2040–2068, 2022. PMID 34968080.
- [48] Toni Schmader, Jessica Whitehead, and Vicki H. Wyatt. A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants. *Sex Roles*, 57(7–8):509–514, 2007.
- [49] Frank L. Schmidt and John E. Hunter. The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2):262–274, 1998.
- [50] Steven E. Scullen, Michael K. Mount, and Maynard Goff. Understanding the latent structure of job performance ratings. *Journal of Applied Psychology*, 85(6):956–970, 2000.
- [51] Michael Spence. Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374, 1973.
- [52] Floris Takens. Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381. Springer, 1981.
- [53] Paul J. Taylor, Karl Pajo, Gordon W. Cheung, and Patrick Stringfield. Dimensionality and validity of a structured telephone reference check procedure. *Personnel Psychology*, 57:745–772, 2004.
- [54] Philip E. Tetlock. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press, 2005.
- [55] Philip E. Tetlock and Dan Gardner. *Superforecasting: The Art and Science of Prediction*. Crown, New York, 2015.
- [56] Frances Trix and Carolyn Psenka. Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse and Society*, 14(2):191–220, 2003.
- [57] Susan van Rooyen, Fiona Godlee, Stephen Evans, Richard Smith, and Nick Black. Effect of blinding and unmasking on the quality of peer review: A randomized trial. *JAMA*, 1998. PMID 9676666.
- [58] Chockalingam Viswesvaran, Deniz S. Ones, and Frank L. Schmidt. Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81:557–574, 1996.
- [59] Chockalingam Viswesvaran, Frank L. Schmidt, and Deniz S. Ones. Is there a general factor in ratings of job performance? a meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology*, 90(1):108–131, 2005. PMID 15641893.
- [60] Lawrence J. White. Markets: The credit rating agencies. *Journal of Economic Perspectives*, 24(2):211–226, 2010.
- [61] Lawrence J. White. Credit rating agencies: An overview. *Annual Review of Financial Economics*, 5:593–615, 2013.

- [62] William C. Wimsatt. Robustness, reliability, and overdetermination. In Marilyn B. Brewer and Barry E. Collins, editors, *Scientific Inquiry and the Social Sciences*, pages 124–163. Jossey-Bass, San Francisco, 1981. Reprinted in Wimsatt, *Re-Engineering Philosophy for Limited Beings*, Harvard University Press, 2007, and in Soler et al. (eds.), *Characterizing the Robustness of Science*, Springer, 2012, pp. 61–78, DOI 10.1007/978-94-007-2759-5_2.
- [63] Lynne G. Zucker. Production of trust: Institutional sources of economic structure, 1840–1920. *Research in Organizational Behavior*, 8:53–111, 1986.