

EPISTEMIC INJUSTICE IN ALGORITHMIC DECISION-MAKING IN SOCIAL PROTECTION SYSTEMS: OPERATIONALISING FRICKER'S TESTIMONIAL AND HERMENEUTICAL INJUSTICE FOR AUTOMATED CASE-MANAGEMENT

Reda Mansouri 

University of Algiers 2 “Abou El Kacem Saâdallah”
Algiers, People's Democratic Republic of Algeria
E-mail: reda.mansouri@univ-alger2.dz

Primljeno: 02.08.2025. Prihvaćeno: 17.10.2025.

Original Scientific Article

DOI: <https://doi.org/10.65932/CR-2025-2-5> UDC: 1 Fricker M.: [165:17]

Abstract: The article addresses a precise question that arises at the intersection of Miranda Fricker's theory of epistemic injustice and the ongoing automation of social-protection case-management systems across Europe, the United Kingdom, and Australia: does the user lose her epistemic standing when she is categorised by an algorithm whose interpretive horizon she does not share, and whose categorical apparatus she cannot inspect? Drawing on Fricker's foundational distinction between testimonial injustice (a credibility deficit assigned to a speaker on the basis of identity-prejudice) and hermeneutical injustice (a structural deficit in shared interpretive resources that disadvantages a marginal group in articulating its experience), the article develops a typology of five modes of algorithmic epistemic closure through which automated case-management systems systematically reproduce both forms of injustice. The five modes are: (1) categorical translation, by which the user's narrative is compressed into pre-defined system fields that admit only of the kinds of testimony the algorithm can recognise; (2) calibration silencing, by which an algorithm trained on biased historical data systematically discounts certain classes of testimony as statistically less credible; (3) opacity, by which the black-boxed character of the model deprives the user of the interpretive resources required to make sense of the decision; (4) temporal compression, by which the speed of automated processing forecloses the space for nuance, contextualisation, and follow-up that human caseworkers would otherwise provide; and (5) appeal asymmetry, by which formal redress procedures presuppose a level of algorithmic literacy and infrastructural access that the affected user characteristically lacks. The typology is applied empirically to three documented cases — the Dutch SyRI welfare-fraud detection system, the United Kingdom's 2020 Ofqual A-level grading algorithm, and the Australian Robodebt automated debt-recovery scheme — drawing on the peer-reviewed analyses of these cases by van Bekkum and Borgesius (2021), Smith (2020), and Nikidehaghani, Andrew and Cortese (2023). The article argues that the five modes do not act in isolation; they form an integrated apparatus whose cumulative effect is the conversion of a person seeking social support into what I term an “algorithmically illegible subject” — a position from which neither testimony nor self-interpretation can be successfully advanced within the system as it stands.

Keywords: *epistemic injustice, Miranda Fricker, testimonial injustice, hermeneutical injustice, algorithmic decision-making, social protection, welfare automation, SyRI, Ofqual, Robodebt, algorithmic epistemic closure, algorithmically illegible subject.*

INTRODUCTION

When Miranda Fricker formulated, in *Epistemic Injustice: Power and the Ethics of Knowing* (Fricker, 2007), the distinction between testimonial and hermeneutical injustice, the social technologies under analysis were predominantly face-to-face — workplace interactions, courtroom testimony, casual exchanges in which a hearer assigns more or less credibility to a speaker on the basis of an identity-prejudice the hearer holds. Fricker's analytic apparatus was designed for, and sharpened by, situations in which a speaker tries to be heard and a hearer fails to hear adequately. Two decades later, however, a substantial fraction of the encounters that determine whether a person retains access to social-protection benefits, qualifies for housing, receives a fair examination grade, or has a debt notice raised against her, no longer take the form of a face-to-face exchange at all. They take the form of an algorithmic processing event: a set of input fields, a model trained on historical data, a categorical output, and an appeal pathway whose architecture mirrors the model's own logic. The question this article addresses is whether Fricker's framework still works when the hearer is no longer a person, and what — if it does work — its operationalisation actually looks like.

The empirical material is, by 2024 and 2025, sufficiently dense to permit serious testing. The Dutch System Risico Indicatie (SyRI), a welfare-fraud detection system halted by the District Court of The Hague on 5 February 2020 for violation of Article 8 of the European Convention on Human Rights, has been the subject of several peer-reviewed legal-ethical analyses, of which van Bekkum and Borgesius (2021) is the most cited. The United Kingdom's Ofqual A-level grading algorithm, deployed in August 2020 in lieu of cancelled examinations, withdrawn within five days under public pressure, and analysed in the peer-reviewed literature most prominently by Smith (2020), illustrates with unusual sharpness how a single algorithmic event can produce a population-scale instance of what Fricker would recognise as testimonial injustice. The Australian Robodebt scheme, in operation between 2016 and 2019 and the subject of a Royal Commission whose report was delivered in July 2023, has generated a peer-reviewed corpus including Nikidehaghani, Andrew and Cortese (2023) in *Accounting, Auditing & Accountability Journal* and Chowdhury (2024) in the *Australian Journal of Human Rights*, both of which treat the scheme as a structural rather than incidental failure. To these three cases the present article adds a series of conceptual references to the recent philosophical literature on algorithmic and data-science forms of epistemic injustice — most centrally Symons and Alvarado (2022) in *Synthese* and Milano and Prunkl (2025) in *Philosophical Studies* — and draws on the empirical work of van Toorn and Scully (2024) in *Disability & Society* on the National Disability Insurance Scheme in Australia.

The central research question of this article is the following: by what specific structural mechanisms do automated case-management systems convert the testimony and self-interpretation of an affected user into a position from which neither can be successfully advanced, and how does the resulting “algorithmically illegible subject” map onto Fricker's distinction between testimonial and hermeneutical injustice? Three hypotheses follow. First, the production of epistemic injustice by algorithmic case-management systems is not a residual failure of design but a structural feature of how such systems classify, decide, and explain. Second, this structural feature can be analysed in terms of a finite, recurrent set of modes of “epistemic closure”, each of which can be located in concrete operational practices that auditors, advocates, and affected users can identify. Third, these modes do not collapse into one another; they form a typology in which some modes preferentially produce testimonial injustice (a credibility deficit assigned to the speaker), some preferentially produce hermeneutical injustice (a deficit of shared interpretive resources), and some produce both at once.

The original contribution of this article is the construction of a typology of *five modes of algorithmic epistemic closure*: categorical translation, calibration silencing, opacity, temporal compression, and appeal asymmetry. The typology is, to the best of my reading of the literature, the first systematic operationalisation of Fricker's distinction for the specific empirical setting of automated case-management in social-protection systems, applied across three documented cross-jurisdictional cases. The typology is heuristic as well as descriptive: it is intended both to organise the existing empirical record and to direct future empirical investigation toward the specific mechanisms by which epistemic injustice is produced.

The article proceeds as follows. The next section develops the literature review and methodology. The section after presents the typology and its empirical application to SyRI, Ofqual, and Robodebt. Four analytical sections then explore the theoretical implications: first, the productivity of Fricker's framework for the algorithmic setting; second, the systematic anatomy of the five modes; third, the comparative analysis across the three cases; and fourth, the conceptual stakes of the “algorithmically illegible subject” as a position that both Fricker's testimonial and hermeneutical injustices converge upon. The conclusion revisits the hypotheses, articulates the original contribution, and outlines policy and research implications.

LITERATURE REVIEW AND METHODOLOGY

Literature Review

Fricker's *Epistemic Injustice* (Fricker, 2007) introduced two analytic categories that have since become foundational for any work on epistemic harms. Testimonial injustice occurs when a speaker is given less credibility than she is due because of an identity-prejudice held by the hearer; the canonical example is the woman whose claim about a male colleague's behaviour is discounted because of a sexist prior. Hermeneutical injustice occurs when a person, characteristically a member of a marginalised group, lacks the conceptual or linguistic resources to make sense of her own social experience because the surrounding interpretive community has, structurally, not produced those resources; the canonical example is the woman who, before the term “sexual harassment” entered the public lexicon, lacked the vocabulary to articulate what was happening to her. The two forms of injustice differ in their target — testimonial injustice is local, occurring in a specific exchange, while hermeneutical injustice is structural, occurring at the level of shared interpretive resources — but Fricker treats them as complementary, with hermeneutical injustice often providing the underlying conditions that make individual testimonial injustices possible.

The reception of Fricker's framework in philosophy of technology, after a long delay, has been productive. Symons and Alvarado (2022), in *Synthese*, argue that data-science technologies are systematically liable to produce both testimonial and hermeneutical harms, and propose a conceptual taxonomy of how such harms differ from more familiar forms of harm produced by such systems (privacy violation, discrimination, autonomy interference). Their argument is that epistemic harms are conceptually distinguishable from these familiar harms even when they intersect with them in practice, and that this distinction matters because the remediation strategies for epistemic harms (restoration of standing, expansion of interpretive resources) differ from those for the familiar harms (restitution, anti-discrimination law, consent regimes). Milano and Prunkl (2025), in *Philosophical Studies*, develop a complementary line, focusing specifically on hermeneutical injustice as it arises through algorithmic profiling. They identify a novel mechanism, which they call “epistemic fragmentation”, by which algorithmic categorisation depletes the conceptual resources that would otherwise be available for a person to interpret and evaluate her own experiences within a profiled category. Both papers position their analysis as

extensions of Fricker rather than departures from her — the philosophical commitment to her framework remains intact, but the empirical setting is transformed.

The empirical literature on specific automated case-management systems has, in the last six years, become substantially richer. The Dutch SyRI case is the locus classicus. Van Bekkum and Borgesius (2021), in the *European Journal of Social Security*, provide the canonical legal-ethical analysis of the District Court's February 2020 judgment, focusing on the court's reasoning that SyRI legislation contained insufficient safeguards against discriminatory and stereotyping effects and that its operation could not be rendered compatible with Article 8 of the European Convention on Human Rights. Their analysis demonstrates that the court located the violation not at the level of a particular bad outcome but at the level of the system's structural opacity and its incapacity to render its risk-scoring rationale transparent to affected individuals. From the perspective of Fricker's framework, the court's reasoning aligns with what one would expect a structural-epistemic analysis to produce: the harm was hermeneutical before it was testimonial, in the sense that affected users were deprived of the very resources required to interpret their own classification.

The Ofqual case is the most temporally compressed of the three. Smith (2020), in *AI & Society*, provides one of the earliest peer-reviewed analyses, focusing on the algorithmic bias that resulted in the systematic downgrading of A-level results for students from disadvantaged schools and the systematic upgrading of results for students from elite private schools. The structural mechanism, as Smith documents, was the algorithm's heavy reliance on each school's historical performance, which produced a regression-to-the-school-mean effect that penalised individual students who would have outperformed their school's prior cohort. From the perspective of this article's typology, the Ofqual case is paradigmatic of what I will describe as calibration silencing: the algorithm could not “hear” the testimony of a student whose individual performance would have departed from her school's historical baseline, because such a testimony had no representation in the model's input fields.

The Australian Robodebt scheme is the most extensively documented of the three in the academic literature, in part because the Royal Commission report of July 2023 generated a wave of secondary analysis. Nikidehaghani, Andrew and Cortese (2023), in *Accounting, Auditing & Accountability Journal*, treat Robodebt through the lens of “algorithmic accountability” — the question of how algorithmic systems are made (or fail to be made) answerable to those they affect. Their analysis foregrounds the role of accounting techniques in masking the political-economic commitments of the scheme: by reframing welfare recipients as “potential cheats” through the categorical apparatus of automated debt detection, the system simultaneously discounted their testimony (a testimonial-injustice operation) and depleted the interpretive resources required to contest the categorisation (a hermeneutical-injustice operation). Chowdhury (2024), in the *Australian Journal of Human Rights*, complements this analysis by reading the same scheme through international human-rights frameworks (ICCPR, ICESCR), and shows that the scheme produced violations of administrative-law standards that map closely onto what Fricker would identify as testimonial harm: the systematic refusal to accept the user's own account of her income and circumstances unless that account fitted the algorithm's pre-defined input fields.

A related and growing strand of empirical literature concerns the reception side: how affected users perceive and respond to algorithmic decision-making. Starke, Baleis, Keller and Marcinkowski (2022), in *Big Data & Society*, provide a systematic review of the empirical literature on fairness perceptions of algorithmic decision-making, drawing on fifty-eight studies. Their review documents that perceptions of algorithmic unfairness are widespread, that they correlate with perceptions of opacity, and that they are exacerbated when affected users have no recourse

to a human decision-maker. Van Toorn and Scully (2024), in *Disability & Society*, provide a complementary qualitative analysis of the National Disability Insurance Scheme in Australia, demonstrating that disabled people interacting with algorithmic eligibility-determination systems develop their own interpretive frameworks for what is happening to them, even when the official system does not provide such frameworks. Their analysis is particularly valuable for the present argument because it documents an empirical instance of what Fricker would call hermeneutical resistance: affected users build interpretive resources outside the official system to compensate for those they are denied within it.

Two further bodies of literature inform the methodological frame. The first is the comparative literature on algorithmic decision-making in public administration, which documents the cross-jurisdictional pattern by which similar epistemic harms recur in formally distinct legal systems (Symons & Alvarado, 2022; Starke et al., 2022). The second is the now-substantial literature on opacity and explainability in algorithmic systems, which provides the technical-epistemic background for understanding why hermeneutical injustice in this domain is so persistent: even where remediation is sought through “explainable AI” or post-hoc explanation tools, the explanations themselves often fail to provide the interpretive resources required for the affected user to make sense of her own classification (Milano & Prunkl, 2025). This article does not enter the technical debate over explainability directly; it relies on the literature's documented finding that the resources currently available are insufficient for the form of interpretive understanding that Fricker's framework requires.

A clear research gap remains. Although the literature has, since 2019, increasingly applied Fricker's framework to algorithmic systems, there is no systematic operationalisation of testimonial and hermeneutical injustice for the specific setting of automated case-management in social protection, with comparative empirical application across documented cases. This article fills that gap.

Research Methodology

The methodology combines theoretical analysis, typology construction, and comparative case-study application. The first methodological component is a critical synthesis of Fricker's *Epistemic Injustice* (Fricker, 2007), with particular attention to the conceptual distinction between testimonial and hermeneutical injustice and to the conditions under which Fricker's analysis presupposes a face-to-face hearer. Rather than restating Fricker's argument in full, I extract the operational elements of her framework — credibility assignment, identity-prejudice, hermeneutical marginalisation, conceptual gap — and treat each as a feature whose translation to the algorithmic setting may or may not preserve its analytic content.

The second methodological component is the construction of the typology of five modes of algorithmic epistemic closure. The typology is derived through a two-way procedure. Deductively, I begin from Fricker's two categories and ask: what structural features of automated case-management systems are most likely to produce instances of each? Testimonial injustice presupposes a hearer who systematically discounts a speaker; in algorithmic systems, the analogous mechanism is the model that systematically discounts certain inputs. Hermeneutical injustice presupposes a structural deficit in shared interpretive resources; in algorithmic systems, the analogous mechanism is opacity and the absence of conceptual resources for the user to interpret her own classification. Inductively, I survey the empirical literature on SyRI, Ofqual, Robodebt, and adjacent cases (Allhutter et al. on the Austrian AMAS system; van Toorn and Scully on the NDIS) and identify the recurrent operational practices that have been documented as producing epistemic harm. Five modes emerge from the convergence of these two lines:

categorical translation, calibration silencing, opacity, temporal compression, and appeal asymmetry.

The third methodological component is the comparative application of the typology to three documented cases. The three cases are selected for variation along several dimensions. SyRI is a fraud-detection system; Ofqual is a grade-determination system; Robodebt is a debt-recovery system. SyRI was halted by judicial review; Ofqual was withdrawn under political pressure; Robodebt operated for over three years before being legally invalidated. SyRI and Ofqual are European; Robodebt is Australian. This variation allows the typology to be stress-tested across different institutional, legal, and operational contexts. For each case, I apply the five modes systematically, using the peer-reviewed analyses (van Bekkum & Borgesius, 2021; Smith, 2020; Nikidehaghani et al., 2023; Chowdhury, 2024) as the primary empirical record.

The fourth methodological component is the development of the concept of the “algorithmically illegible subject” — the position produced by the cumulative operation of the five modes, in which neither testimony nor self-interpretation can be successfully advanced within the system as it stands. This concept is, in part, descriptive (it names a position that the empirical record documents); in part, theoretical (it specifies a structural relation between Fricker's two forms of injustice and the algorithmic setting); and in part, evaluative (it implies a normative claim about the conditions under which automated case-management systems can be ethically defended). The concept is intended as a complement to Fricker's framework, not a replacement of it: the algorithmically illegible subject is the contemporary form that the victim of compound testimonial-and-hermeneutical injustice takes when her interlocutor is no longer a human hearer.

The fifth methodological component is data-source triangulation. For each case, I rely on at least two peer-reviewed analyses where available; where peer-reviewed analyses converge, I treat the converged finding as well-supported. For the Dutch SyRI case, the primary source is van Bekkum and Borgesius (2021), supplemented by Symons and Alvarado (2022) for the philosophical-conceptual frame. For the Ofqual case, the primary source is Smith (2020), supplemented by Starke et al. (2022) for the broader fairness-perception literature. For the Robodebt case, the primary sources are Nikidehaghani et al. (2023) and Chowdhury (2024), supplemented by van Toorn and Scully (2024) for the parallel disability-rights case. Conceptual claims are anchored in Fricker (2007), Symons and Alvarado (2022), and Milano and Prunkl (2025).

The methodological limits of the study are four. First, I do not have direct access to internal documentation from the agencies that operated the three systems, and rely on the publicly available record as filtered through peer-reviewed academic analyses. The peer-reviewed analyses themselves rely on judicial findings, parliamentary inquiries, and freedom-of-information disclosures, but a fully internal-perspective analysis would be richer. Second, Fricker's framework is the conceptual frame used; alternative philosophical frameworks (Mbembean necropolitics, Foucauldian governmentality, the philosophy of technology of Don Ihde) might generate different categories. The choice of Fricker is justified by the centrality of her framework to the existing literature on epistemic injustice in algorithmic systems, but other choices remain defensible. Third, the typology is generated in part inductively from documented cases; its applicability to undocumented or differently structured cases (for example, Chinese social-credit systems, or African mobile-money welfare systems) is a matter for future research and is not claimed in advance. Fourth, the boundary between the five modes is not always sharp in practice; some empirical instances belong to two or more modes simultaneously, and the typology is best understood as a set of analytically separable features rather than as a partition of empirical instances into mutually exclusive classes.

RESEARCH RESULTS

The empirical analysis of the three documented cases yielded findings that organise into three blocks. The first block presents the typology of the five modes of algorithmic epistemic closure with their conceptual definitions and primary empirical loci. The second block applies the typology to SyRI, Ofqual, and Robodebt and documents the modes operating in each case. The third block characterises the cumulative effect of the modes through the concept of the “algorithmically illegible subject”.

Table 1 presents the typology. The first mode, *categorical translation*, names the structural mechanism by which the user's narrative is compressed into pre-defined system input fields. In SyRI, the input categorisation reduced a citizen's life situation to a vector of indicators (employment status, residence in a designated postal code, prior receipt of benefits, family composition) that constituted the operational definition of “fraud risk” for the system. In Ofqual, the input compression was even sharper: each student's ranking within her class, her school's historical exam-performance distribution, and a small number of demographic and contextual fields constituted the entirety of the input. In Robodebt, income testimony was systematically replaced by a calculated “average fortnightly income” derived from annual taxation data — a translation that, as Nikidehaghani et al. (2023:684) document, mathematically converted a population of irregular-earnings welfare recipients into a population of putative debtors. From Fricker's perspective, categorical translation is primarily a producer of testimonial injustice: it systematically structures the system's “hearing” so that only certain forms of testimony register as relevant.

The second mode, *calibration silencing*, names the structural mechanism by which an algorithm trained on biased historical data systematically discounts certain classes of testimony as statistically less credible. The Ofqual case is paradigmatic: the algorithm's reliance on each school's prior-cohort performance distribution meant that a high-performing student in a low-performing school could not have her individual performance “heard” by the model except as an outlier whose ranking was compressed back toward the school's historical mean (Smith, 2020:1077). The SyRI case, although less mathematically transparent, exhibited a parallel mechanism: the model was trained on data from prior fraud investigations that themselves reflected historical patterns of investigative attention — patterns that, as the District Court of The Hague noted in its February 2020 judgment, had a documented tendency to concentrate on lower-income postal-code districts (van Bekkum & Borgesius, 2021:330). The Robodebt case showed calibration silencing through the absence of any mechanism by which a recipient's testimony of irregular employment patterns could be admitted as a corrective to the averaged-income calculation. From Fricker's perspective, calibration silencing is the algorithmic analogue of testimonial injustice rooted in identity-prejudice: the model's parameters embed historical patterns of credibility-allocation, and the testimony of a person whose identity correlates with the historically discounted patterns is, as a structural matter, less likely to “register”.

The third mode, *opacity*, names the structural mechanism by which the black-boxed character of the model deprives the user of the interpretive resources required to make sense of the decision. In all three cases, the affected users could not, with the resources available to them, construct an explanation of why the system had reached the decision it had. SyRI was, by design, kept opaque — the District Court of The Hague specifically cited the system's opacity as one of the grounds for its judgment that Article 8 of the European Convention on Human Rights had been violated (van Bekkum & Borgesius, 2021:331). Ofqual published its model after deployment, but the published documentation was dense and required statistical training to interpret; for the typical affected student or her parents, the model remained operationally

opaque. Robodebt was the most strongly opaque of the three: the basis on which a debt notice had been calculated was, in many cases, not communicated to the recipient at all, and the system was structurally designed to place the burden of interpretation on the recipient (Nikidehaghani et al., 2023:687). From Fricker's perspective, opacity is the primary producer of hermeneutical injustice in the algorithmic setting: it is precisely the gap in shared interpretive resources that prevents the user from making sense of her own classification.

The fourth mode, *temporal compression*, names the structural mechanism by which the speed of automated processing forecloses the space for nuance, contextualisation, and follow-up. In SyRI, risk-scoring was performed at the system level on a periodic basis, so that affected individuals received an investigation-trigger before any human caseworker had reviewed the case in detail. In Ofqual, the entire algorithmic process from input ingestion to grade output was completed in a matter of days during August 2020, with no opportunity for individual students to contextualise their results before the grades were officially issued. In Robodebt, debt notices were issued automatically on the basis of the averaged-income calculation, again without prior human review of the individual case (Chowdhury, 2024:28). The recipient was placed in the position of having to react to a decision rather than participate in the construction of it. From Fricker's perspective, temporal compression is a hybrid producer: it limits both testimonial injustice (the user does not have time to advance her testimony) and hermeneutical injustice (she does not have time to construct an interpretation of her own situation).

The fifth mode, *appeal asymmetry*, names the structural mechanism by which formal redress procedures presuppose a level of algorithmic literacy and infrastructural access that the affected user characteristically lacks. In all three cases, an appeal pathway formally existed; in all three cases, the burden of using it was systematically disproportionate to the resources available to a typical affected user. SyRI investigation-triggers could in principle be challenged, but the mechanisms required users to know that their data had been processed and to engage with administrative-law procedures whose architecture mirrored the system's own logic (van Bekkum & Borgesius, 2021:333). Ofqual offered an appeal mechanism in the days following the algorithm's deployment, but the mechanism was overwhelmed and the criteria for successful appeal were unclear (Smith, 2020:1078). Robodebt offered a debt-review mechanism that required the recipient to assemble employment records spanning the period of the disputed debt — a burden that, as Chowdhury (2024:31) documents, fell disproportionately on those least able to bear it. From Fricker's perspective, appeal asymmetry compounds both testimonial and hermeneutical injustices: it formally provides an avenue for testimony to be reasserted, but materially constructs that avenue so as to be most difficult to use precisely for those most likely to need it.

Mode	Definition	Primary empirical locus	Fricker mapping
Categorical translation	Compression of narrative into pre-defined input fields	Robodebt averaged-income calculation	Testimonial (primary)
Calibration	Algorithm trained on biased data	Ofqual school-baseline regression; SyRI postal-code priors	Testimonial (primary)
Silencing	discounts certain testimony	SyRI judicial finding; Ofqual; Robodebt non-explained	Hermeneutical (primary)
Opacity	Black-boxed model deprives user of interpretive resources	Ofqual days-long pipeline; Robodebt automatic notices	Compound (testimonial + hermeneutical)
Temporal compression	Speed forecloses nuance and contextualisation	SyRI admin-law thresholds; Ofqual overwhelmed appeal; Robodebt records-burden	Compound (testimonial + hermeneutical)
Appeal asymmetry	Redress procedures presuppose algorithmic literacy		

Table 1. Five modes of algorithmic epistemic closure with primary empirical loci and Fricker mapping
 Sources: Fricker (2007); van Bekkum and Borgesius (2021); Smith (2020); Nikidehaghani, Andrew and Cortese (2023); Chowdhury (2024); Symons and Alvarado (2022); Milano and Prunkl (2025); Starke et al. (2022); van Toorn and Scully (2024).

The second block of results presents the comparative application of the typology to the three cases. SyRI exhibited all five modes, with opacity (mode 3) and appeal asymmetry (mode 5) being judicially salient, in the sense that the District Court of The Hague's reasoning explicitly identified the structural-opacity and the inadequacy of the safeguards as the operative grounds for invalidation (van Bekkum & Borgesius, 2021:332). The case is striking because the modes were diagnosed not by an academic ex-post analysis but by the court itself in its February 2020 judgment, which means that the typology offered here finds an unusual degree of judicial corroboration. Ofqual exhibited all five modes, with calibration silencing (mode 2) being the most analytically distinctive, because the regression-to-the-school-mean structure that Smith (2020) documents constitutes a textbook case of an algorithm "hearing" certain testimony only insofar as it conformed to a historically biased baseline. Robodebt exhibited all five modes with particular intensity at modes 1, 4, and 5 — the categorical translation of irregular income into averaged income, the temporal compression that produced automatic debt notices without human review, and the appeal asymmetry that placed the records-burden on the recipient. The fact that all three cases exhibited all five modes, with no single mode operating in isolation, supports the second hypothesis of the article: epistemic injustice in algorithmic case-management is produced through the cumulative operation of an integrated apparatus, not through individual design failures.

The third block of results presents the cumulative effect through the concept of the algorithmically illegible subject. By the time the five modes have run their course in any of the three cases, the affected user is positioned such that her testimony cannot effectively register (modes 1, 2, 4 cumulatively); her self-interpretation cannot effectively be developed (mode 3); and her appeal cannot effectively be advanced (modes 4, 5 cumulatively). This is the position I have called the algorithmically illegible subject. The position is not equivalent to having no rights — formal rights remain — and it is not equivalent to having no testimony — testimony continues to be offered. It is, more precisely, a structural position from which neither rights nor testimony can be successfully advanced within the system as it stands. The three cases differ in the speed with which the position is produced (Ofqual produced it in days; Robodebt over years), in the demographic composition of those most affected (Ofqual primarily affected school-leaving teenagers; SyRI primarily affected lower-income postal-code residents; Robodebt primarily affected welfare recipients with irregular employment), and in the eventual route by which the position was challenged (judicial review for SyRI; political withdrawal for Ofqual; royal commission for Robodebt). Despite these differences, the structural position itself is recognisably identical across the three cases.

THE PRODUCTIVITY OF FRICKER'S FRAMEWORK FOR THE ALGORITHMIC SETTING

Before turning to the systematic anatomy of the five modes, I want to address a question that the empirical findings have already implicitly raised: does Fricker's framework, originally developed for face-to-face epistemic encounters, actually translate to the algorithmic setting? My argument in this section is that it does, but only if certain features of her framework are extracted analytically rather than carried over wholesale, and that the extracted features turn out to be among the most theoretically robust elements of her account.

The element of Fricker's framework that translates most directly is the distinction between testimonial and hermeneutical injustice itself. In the face-to-face setting, testimonial injustice is a credibility deficit assigned by a specific hearer to a specific speaker; in the algorithmic setting, it is a credibility deficit structurally embedded in the model's parameters and input architecture, applied at population scale. The form of the harm — the speaker's testimony is given less weight

than it is due — is preserved; what changes is the agent of the discounting. Whether one regards this as a translation of Fricker's framework or as an extension of it depends on whether one regards “agency” as a necessary condition for testimonial injustice. Fricker's own treatment, particularly in her later writings, suggests that the structural production of credibility deficits at population scale is recognisably continuous with her original case (Symons & Alvarado, 2022:13).

Hermeneutical injustice translates more straightforwardly still. The original concept names a structural deficit in shared interpretive resources that disadvantages a marginalised group in articulating its experience; the algorithmic setting produces precisely this kind of deficit when the interpretive resources required to make sense of one's own algorithmic classification are not made available to those classified. Milano and Prunkl (2025), in their *Philosophical Studies* paper, identify this as “epistemic fragmentation”: the algorithmic apparatus does not merely fail to provide interpretive resources; it actively fragments the shared resources that would otherwise be available, by classifying users into categories whose conceptual content is opaque even to other users in the same category. This extension is, by my reading, faithful to Fricker's original framework: the harm Fricker named in 2007 has a structural producer she did not anticipate, but the structure of the harm itself is preserved.

What does not translate as directly is Fricker's concept of identity-prejudice as the source of testimonial injustice. In the face-to-face setting, identity-prejudice is a psychological state held by the hearer that produces the credibility deficit; in the algorithmic setting, the analogous source is the historical bias embedded in training data, which is not psychological in any direct sense. Symons and Alvarado (2022:18) argue that this is a difference of mechanism, not of structure: the historical-data bias produces credibility deficits in patterns that match the patterns produced by individual identity-prejudice, even though the immediate cause is different. This is, I think, the right analysis. The pattern is preserved across the change of mechanism, and the moral significance of the harm is preserved with it.

A further element of Fricker's framework that translates productively is her account of the conditions under which hermeneutical injustice is most likely to occur. Fricker argues that hermeneutical injustice tends to arise where a marginalised group has been systematically excluded from the production of shared interpretive resources, so that its experiences are systematically under-represented in the available conceptual vocabulary. Algorithmic systems exhibit a precise analogue: training data is, by construction, a record of past administrative decisions, and those decisions reflect historical patterns of attention and inattention. A group that has historically been attended to in a particular way will have its experiences encoded into the model's interpretive horizon; a group that has not been so attended to will not. The predictable result, as Eubanks (2018) documents at length for US welfare contexts and as Nikidehaghani et al. (2023) document for the Australian Robodebt case, is that the hermeneutical resources of the model — what it is able to recognise as a kind of experience worth registering — track the historical patterns of administrative attention rather than the actual distribution of human experience.

Several scholars have raised concerns about the limits of Fricker's framework in this setting, and these concerns merit acknowledgement. Symons and Alvarado (2022:21) note that Fricker's framework presupposes a target individual who is morally salient as a knower, whereas algorithmic systems frequently produce harm at the population level in ways that may not be reducible to harms to individual knowers. This is a serious worry, but I think it is a worry about how the harm is to be measured and remediated rather than about whether the harm exists. The cumulative effect of the five modes of algorithmic epistemic closure is to produce a structural position from which individual users cannot successfully advance testimony or self-interpretation; the harm to each individual is real, even if the production of the harm is structural.

Milano and Prunkl (2025), responding to a similar worry, develop their concept of epistemic fragmentation precisely as a way of preserving the individual-level moral salience while acknowledging the structural mechanism.

The more difficult question is whether Fricker's framework remains the best framework, or whether other available frameworks (relational ethics, virtue epistemology, feminist standpoint theory, the philosophy of technology) might be more productive. My provisional view is that Fricker's framework remains the most precisely calibrated to the kind of harm at issue, because it directly names the relation between credibility, interpretive resources, and structural disadvantage. Alternative frameworks — relational ethics, in particular — provide complementary but not substitutive resources. The argument of this article does not depend on the claim that Fricker's framework is uniquely correct; it depends only on the weaker claim that it is one productive framework, and that its productivity in the algorithmic setting is sufficient to motivate the typology developed here.

THE ANATOMY OF THE FIVE MODES

This section develops the systematic anatomy of the five modes of algorithmic epistemic closure, paying particular attention to the operational details that distinguish each mode and to the empirical record that documents each in the three cases.

Categorical translation, as the first mode, operates at the input layer of an algorithmic system. The user's situation is, in principle, an open-ended narrative; the system requires that situation to be expressed as a tuple of values across pre-defined fields. The translation is unavoidable as a technical matter — algorithms operate over structured input, not over open-ended narrative — but the choice of input fields encodes a substantive theory of what is and is not relevant to the decision. In Robodebt, the choice to translate “income” as a single annual figure averaged across fortnights encoded a theory of welfare receipt that was empirically false for any recipient with irregular employment, but the encoding produced an authoritative output regardless. As Nikidehaghani et al. (2023:684) document, the resulting categorical apparatus was central to the reframing of welfare recipients as “potential cheats” whose individual narratives had no place in the system's operational logic. The user's testimony of irregular employment had no input field in which to be entered, and so was, in the system's terms, structurally unutterable. From Fricker's perspective, this is testimonial injustice in its most operational form: not that the system “hears” the testimony and discounts it, but that the system's input architecture provides no place for the testimony to be advanced.

Calibration silencing, as the second mode, operates at the model layer. Once input has been translated, the model assigns weights to different combinations of input values, and those weights are calibrated against historical training data. Where the training data reflects historical patterns of bias — disproportionate investigative attention, systematically lower historical achievement in particular schools, patterns of debt that themselves reflect prior automated decisions — the calibration encodes those patterns into the model's operational behaviour. Smith (2020:1078) documents the mechanism for the Ofqual case with particular clarity: the algorithm's reliance on each school's historical performance distribution meant that an individual student's “true” performance, in the sense of the performance she would have achieved had the examination been held, was systematically compressed toward her school's historical mean. A high-performing student in a historically low-performing school was, in the system's terms, statistically less credible than the same student would have been in a historically high-performing school. The mechanism is not psychological identity-prejudice on the part of any individual hearer; it is structural, statistical, and operative across millions of decisions simultaneously.

Opacity, as the third mode, operates at the output-explanation layer. Once the model has produced a decision, the affected user is in principle entitled to understand on what basis the decision was reached. In all three cases, this entitlement was either formally absent or materially unfulfilled. The District Court of The Hague's reasoning in the SyRI judgment singled out opacity as a violation of Article 8 of the European Convention on Human Rights, on the grounds that the inability of affected users to understand the system's risk-scoring rationale constituted a substantive failure of proportionality (van Bekkum & Borgesius, 2021:331). The court's analysis, which is in many ways the most theoretically explicit of the three, identifies opacity as a structural feature of the system rather than a contingent failing of communication. The Ofqual case is more ambivalent: the model was published after deployment, but the published documentation required statistical training to interpret, so that the formal availability of the model did not translate into operational interpretive access for the typical affected user. Robodebt was the most opaque of the three: in many cases the basis of the calculated debt was not communicated to the recipient at all, and recipients reported being unable to determine even which years of taxation data had been used (Chowdhury, 2024:30). From Fricker's perspective, opacity is the primary producer of hermeneutical injustice in the algorithmic setting: the user is denied the very interpretive resources required to make sense of her own classification.

Temporal compression, as the fourth mode, operates at the procedural layer. Automated case-management systems produce decisions on time-scales that are typically faster than human caseworkers would, and this acceleration is often presented as a virtue. From the perspective of the affected user, however, the compression eliminates the procedural pause in which testimony could be advanced, contextualisation provided, and follow-up arranged. Ofqual is the extreme case: the entire pipeline from input ingestion to grade output ran in days, and individual students had no procedural mechanism by which to insert testimony into the pipeline before the output was issued. Robodebt operated on a longer time-scale at the population level — the scheme ran for over three years — but at the individual level, debt notices were issued automatically without prior human review, so that the recipient was placed in the position of having to react to a decision she had had no opportunity to participate in constructing (Chowdhury, 2024:28). In SyRI, the temporal compression operated in a slightly different mode: investigation-triggers were generated periodically, and an affected user might never know that her data had been processed at all. From Fricker's perspective, temporal compression is a hybrid producer: it limits both testimonial injustice (the user has no time to advance her testimony) and hermeneutical injustice (she has no time to construct an interpretation of her own situation). Starke et al. (2022) document, on the basis of their fifty-eight-study systematic review, that perceptions of unfairness in algorithmic decision-making correlate strongly with perceptions of being denied procedural participation — a finding that aligns with the analysis advanced here.

Appeal asymmetry, as the fifth mode, operates at the redress layer. Formal redress mechanisms exist in all three cases, and indeed are required by administrative law in each of the relevant jurisdictions. The structural problem is that the redress mechanisms are themselves built around the same architectural assumptions as the original system. To appeal a SyRI investigation-trigger requires knowing that one has been triggered (which itself requires a level of system-literacy that the system does not provide); to appeal an Ofqual grade requires understanding the algorithmic basis for the grade (which requires interpretive resources that the published documentation does not adequately supply); to appeal a Robodebt notice requires assembling employment records spanning the disputed period (a burden that, as Chowdhury (2024:31) documents, fell most heavily on those whose irregular employment had triggered the calculation in the first place). Van Toorn and Scully (2024:3015) document, in the parallel NDIS case, that the redress procedure presupposes a level of cognitive and infrastructural access that disabled users systematically lack

— and that the redress procedure thus has the structural effect of converting a formal right into a material privilege available only to those who least need it. From Fricker's perspective, appeal asymmetry is a compound producer of testimonial and hermeneutical injustice: it formally provides an avenue for testimony to be reasserted, but materially constructs that avenue so that the very users whose testimony was originally discounted are least able to use it.

The five modes together exhibit a property that I take to be analytically significant: they reinforce one another. Categorical translation produces input architectures that calibrate the model in particular ways; calibration silencing operates within the categorical apparatus that translation has produced; opacity is most damaging where the categorical translation and calibration have already departed from the user's own understanding of her situation; temporal compression eliminates the space in which opacity might be remediated; and appeal asymmetry consummates the closure by ensuring that even formal redress is structurally inaccessible. Each mode considered in isolation is an identifiable source of harm; the cumulative operation of the modes produces the structural position I have called the algorithmically illegible subject.

COMPARATIVE ANALYSIS ACROSS THE THREE CASES

The third analytical section presents a more detailed comparative analysis across the three cases, with attention to what their differences and convergences reveal about the typology and about Fricker's framework.

The Dutch SyRI case is, in some ways, the most theoretically illuminating because it produced an explicit judicial articulation of the harm. The District Court of The Hague's February 2020 judgment, as analysed by van Bekkum and Borgesius (2021), identified opacity and the absence of structural safeguards as the operative grounds for the system's invalidation under Article 8 of the European Convention on Human Rights. From the perspective of this article's typology, the court's reasoning maps onto modes 3 (opacity) and 5 (appeal asymmetry) most directly, but the underlying harm — that affected users could not understand or challenge the basis of their classification — implicates the full set of modes. Van Bekkum and Borgesius (2021:333) note that the judgment was the first in which a court invalidated a welfare-fraud detection system on rights-of-privacy grounds, and that its reasoning has subsequently informed analyses of other European algorithmic welfare systems. The SyRI case thus demonstrates that the structural-epistemic harm produced by the typology's five modes is not only academically describable but legally cognisable.

The Ofqual case, although temporally brief and politically resolved within a week of deployment, is theoretically valuable because it isolates calibration silencing (mode 2) more sharply than the other two cases. The mathematical structure of the Ofqual algorithm — relying heavily on each school's historical performance distribution — produced a population-scale instance of testimonial injustice in which the testimony of individual high-performing students in low-performing schools was systematically discounted. As Smith (2020:1077) documents, this mechanism disproportionately affected students from disadvantaged backgrounds; the algorithm's "calibration" was, statistically, an encoding of educational inequality into operational decision-making. From Fricker's perspective, this is a paradigmatic case of testimonial injustice produced by a non-psychological mechanism — but the form of the harm is identical to the form Fricker originally analysed. The political withdrawal of the algorithm within five days demonstrates that the harm was visible to the affected population almost immediately, which raises an interesting question about why the harm produced by SyRI and Robodebt was not similarly rapidly visible. My provisional answer is that the Ofqual case affected a demographically homogeneous population (school-leaving teenagers across the United Kingdom) whose

collective political voice was unusually effective; SyRI and Robodebt affected populations whose political voice was structurally weaker.

The Australian Robodebt case is the longest-running and most extensively documented of the three, and also the most analytically rich. Nikidehaghani, Andrew and Cortese (2023), in their *Accounting, Auditing & Accountability Journal* analysis, argue that the categorical apparatus of the scheme — the reframing of welfare recipients as “potential cheats” through automated debt calculation — was simultaneously a categorical-translation mechanism (mode 1) and a calibration-silencing mechanism (mode 2), because the operational categories embedded historical assumptions about welfare receipt that themselves reflected biased patterns of attention. Chowdhury (2024), in the *Australian Journal of Human Rights*, extends this analysis through the international human-rights framework, demonstrating that the scheme's structural-epistemic harms produced violations of administrative-law standards that map onto Fricker's testimonial-injustice category. The Royal Commission report of July 2023, although not itself a peer-reviewed academic source, is referenced in the secondary literature as an authoritative finding that the scheme's harms were structural rather than incidental. The Robodebt case is also valuable because it documents the long-run population-level effect of the cumulative operation of the five modes: between 2016 and 2019, approximately 526,000 individuals received debt notices, many of which were subsequently determined to be invalid, and the population-level harms — including documented mental-health impacts and several reported suicides — provide an unusually clear demonstration of the seriousness of the structural-epistemic injustice the typology is designed to identify.

What do the comparative analyses converge on? Three findings, all of which support the article's hypotheses. First, the five modes of algorithmic epistemic closure are not specific to any one of the three cases; they recur across all three, with variations in intensity but not in structure. Second, the modes are not separable in their operation; in each case, they reinforce one another to produce the structural position of the algorithmically illegible subject. Third, the harm produced by the cumulative operation of the modes is morally significant in all three cases, even though the specific consequences (a halted welfare-fraud system, a withdrawn examination algorithm, a Royal Commission into a debt-recovery scheme) differ. The three cases together, taken as a corpus, support the empirical claim that automated case-management systems in social protection systematically produce both testimonial and hermeneutical injustice, and that the production is structural rather than incidental.

A specific point of comparative interest concerns the demographic composition of those most affected. SyRI primarily targeted lower-income postal-code residents (van Bekkum & Borgesius, 2021:328); Ofqual disproportionately affected students from disadvantaged schools (Smith, 2020:1078); Robodebt fell most heavily on welfare recipients with irregular employment patterns (Chowdhury, 2024:25). In all three cases, the population structurally most exposed to the algorithmic apparatus was also the population least equipped to advance the kind of testimony or self-interpretation required to challenge the system's classification. This is not, I think, a coincidence; it reflects the fact that algorithmic case-management is most heavily deployed precisely where administrative cost-pressures are greatest, and administrative cost-pressures are greatest precisely where affected populations have the weakest political voice. From Fricker's perspective, this is a structural feature of the contemporary deployment of automated case-management: the populations most subject to the systems are also those most subject to the conditions under which Fricker's two forms of injustice classically arise.

THE ALGORITHMICALLY ILLEGIBLE SUBJECT

The fourth analytical section develops the conceptual significance of the algorithmically illegible subject as the structural position produced by the cumulative operation of the five modes. The concept is intended to sit alongside, rather than displace, Fricker's existing categories: it names the contemporary form taken by the victim of compound testimonial-and-hermeneutical injustice when her interlocutor is no longer a human hearer.

The algorithmically illegible subject is, formally, a person who has been classified by an automated system in such a way that her testimony cannot effectively register within the system, and her self-interpretation cannot effectively be developed within the system, and her appeal cannot effectively be advanced within the system. Each of these three conditions is produced by a subset of the five modes: testimony is closed by modes 1, 2, and 4; self-interpretation is closed by mode 3; appeal is closed by modes 4 and 5. The conditions are jointly necessary and individually sufficient: a person who satisfies any one condition is in a position of partial illegibility, and a person who satisfies all three is fully illegible to the system. In all three of the cases analysed, the affected users were, by the time the system had run its course, in the fully illegible position.

It is important to distinguish the algorithmically illegible subject from several adjacent concepts. The illegible subject is not Agamben's bare life — she retains formal rights and substantive standing in many other contexts, including most importantly in her ordinary social and familial life; her illegibility is specifically to the algorithmic system. She is not Mbembe's necropolitical subject — the harm at issue is not directly violent in the way necropolitics tracks. She is closer to what Symons and Alvarado (2022:25) call the “epistemic patient” of data-science technologies, but the conceptual content of her position is more specific than that broader category: she is illegible in a precise structural sense that the typology of five modes makes explicit.

The concept has implications both descriptive and normative. Descriptively, it provides a way of naming the convergence of testimonial and hermeneutical injustice in the algorithmic setting — a convergence that Fricker's original framework, written before the ubiquity of automated case-management, did not specifically anticipate but whose existence her framework predicts. Normatively, it implies that the remediation of algorithmic epistemic injustice cannot be limited to the remediation of any single mode in isolation. A system that addresses opacity (mode 3) but leaves categorical translation (mode 1) unchanged will still produce illegible subjects; a system that addresses appeal asymmetry (mode 5) but leaves calibration silencing (mode 2) unchanged will still produce illegible subjects. Effective remediation requires addressing the full set of modes in concert, and the concept of the algorithmically illegible subject names the structural position that any partial remediation will continue to produce.

A further normative implication concerns the relation between the algorithmically illegible subject and the wider social-protection system. In each of the three cases, the affected user was in a relationship of dependency on the system: she required the system's recognition (as eligible for benefits, as fairly graded, as not in debt) in order to access important social goods. The illegibility produced by the cumulative operation of the five modes was therefore not an abstract harm; it was a harm that intersected with, and amplified, the user's prior dependency. This is a structural feature of social-protection contexts that distinguishes them from many other algorithmic-decision contexts (e.g., advertising targeting, content recommendation): the affected user cannot exit the system without losing access to the social goods she depends on. Van Toorn and Scully (2024:3019) develop a parallel point in the disability-rights context, arguing that the

structural dependency of NDIS users on the algorithmic eligibility-determination system means that the user's "consent" to being classified is, in any meaningful sense, structurally pre-empted.

I want to acknowledge a worry that arises here. The concept of the algorithmically illegible subject might be read as totalising — as suggesting that the affected user has no resources of her own with which to resist the system's classification. This would be a misreading. As van Toorn and Scully (2024:3017) document for the NDIS case, affected users frequently develop their own interpretive frameworks for what is happening to them, even when the official system does not provide such frameworks. Such practices of hermeneutical resistance are real and important. The concept of the algorithmically illegible subject does not deny the existence of these resources; it specifies that, within the operation of the system as it stands, those resources cannot effectively register. The user remains a knower; she retains her capacity for self-interpretation; what she lacks is the structural capacity to have either of these recognised by the system that controls her access to the social goods she depends on. This distinction is important because it preserves the moral salience of the affected user's continuing knowing-status while accurately describing the structural position the system places her in.

A final consideration in this section concerns the relation between the algorithmically illegible subject and the political response to the three cases. In each case, the structural-epistemic harm was eventually challenged — through judicial review (SyRI), political withdrawal (Ofqual), or royal commission (Robodebt). The challenges came not from within the system but from outside it: courts, parliaments, civil-society organisations, journalists, affected users mobilised collectively. This pattern is, I think, theoretically important. The algorithmically illegible subject cannot effectively challenge the system from within it (that is part of what it means to be illegible); but external resources — legal, political, social — can be brought to bear on the system and, in the three cases, have been. The political response to the three cases thus illustrates a route of remediation that does not depend on the system itself becoming legible to the affected user, but rather on the broader social-political community providing the resources that the system itself withholds. This is, I take it, what Fricker's framework would predict for the algorithmic setting: hermeneutical injustice is, by its nature, remediated through the expansion of shared interpretive resources, which is necessarily a structural-political achievement rather than an individual one.

CONCLUSION

This article has attempted to demonstrate that Miranda Fricker's distinction between testimonial and hermeneutical injustice can be operationalised for the empirical setting of automated case-management in social-protection systems, through the construction of a typology of five modes of algorithmic epistemic closure. The three hypotheses formulated in the introduction can now be evaluated. The first hypothesis, that the production of epistemic injustice by algorithmic case-management systems is a structural feature rather than a residual failure of design, finds full empirical support in the comparative analysis of the three cases. SyRI, Ofqual, and Robodebt each exhibited all five modes of epistemic closure, and the modes operated cumulatively rather than in isolation. The structural character of the harm is supported by independent peer-reviewed analyses (van Bekkum & Borgesius, 2021; Smith, 2020; Nikidehaghani, Andrew & Cortese, 2023; Chowdhury, 2024) and by the formal findings of the bodies that eventually invalidated each system (the District Court of The Hague, the political withdrawal of Ofqual, the Australian Royal Commission).

The second hypothesis, that the structural feature can be analysed in terms of a finite, recurrent set of modes of epistemic closure, finds support through the typology itself. The five modes were derived through a two-way procedure (deductive from Fricker's framework,

inductive from the empirical record) and were applied across three structurally distinct cases without requiring substantial modification. The convergence of the deductive and inductive sources on the same five modes provides some confidence that the typology has captured something structurally real about the empirical phenomenon, although the possibility of further modes (for example, the economic or psychosocial modes mentioned in the methodology) remains open for future research.

The third hypothesis, that the modes do not collapse into one another but form a typology in which some preferentially produce testimonial injustice, some hermeneutical injustice, and some both, finds support through the Fricker mapping in Table 1. Categorical translation and calibration silencing primarily produce testimonial injustice; opacity primarily produces hermeneutical injustice; temporal compression and appeal asymmetry produce both. This mapping is not exhaustive, and the boundaries between the categories are softer in practice than the table suggests, but the analytical separability of the modes is preserved across the three cases.

The principal original contribution of this article is the typology of five modes of algorithmic epistemic closure, together with the concept of the algorithmically illegible subject as the structural position produced by their cumulative operation. To the best of my reading of the literature, this is the first systematic operationalisation of Fricker's distinction for the specific setting of automated case-management in social-protection systems, applied across three documented cross-jurisdictional cases. The typology is intended to be both descriptive (organising the existing empirical record) and heuristic (directing future empirical investigation toward specific structural mechanisms). The limitations of the study merit honest acknowledgement. First, I have relied on publicly available and peer-reviewed sources rather than on direct access to internal system documentation; this places a ceiling on the granularity of the analysis. Second, Fricker's framework is one productive conceptual lens, but it is not the only available lens; alternative frameworks may yield complementary or competing categorisations. Third, the three cases are all from European or Australian jurisdictions; the typology's applicability to non-Western administrative settings (Chinese social-credit systems, African mobile-money welfare systems, Latin American conditional cash-transfer programmes) is a question for future research. Fourth, the typology is generated in part inductively from the documented cases; further cases may suggest the addition or modification of modes.

Future research should extend the typology to additional documented cases (including the Austrian AMAS system and the National Disability Insurance Scheme), develop a longitudinal version capable of tracking the modes' evolution over time as systems are reformed, and conduct empirical studies of how affected users develop hermeneutical resistance practices outside the official system. The latter line, building on van Toorn and Scully (2024), would be particularly valuable for understanding how the algorithmically illegible subject continues to exercise her knowing-capacity even when the system itself denies her recognition.

Three policy implications follow. First, designers of automated case-management systems for social protection should treat epistemic-injustice risks as structural rather than incidental, and should design systems with explicit attention to all five modes rather than to any single one in isolation. Second, regulatory and oversight bodies should evaluate proposed and deployed systems against the typology, with attention both to the technical features (input architecture, model calibration, explanation provision) and to the procedural features (timeline, appeal pathway). Third, civil-society and advocacy organisations working with affected populations should incorporate explicit attention to algorithmic literacy and to the development of shared interpretive resources, since the empirical record demonstrates that effective remediation depends on the expansion of resources outside the system as much as on changes within it.

A final personal reflection. Writing this article has required me to confront a feature of the current moment that I find genuinely uncomfortable: many of the structural-epistemic harms Fricker named in 2007 are, in 2025, being re-produced at population scale by systems whose operative logics are formally distinct from those Fricker originally analysed but are structurally continuous with them. The conceptual continuity is, in one sense, a vindication of Fricker's framework — it travels well across very different empirical settings. In another sense, it is a depressing finding: the harms Fricker named have not been remediated, only redeployed. The remediation that her framework calls for — the expansion of shared interpretive resources, the equitable distribution of credibility — remains as much a normative task in 2025 as it was in 2007, with the additional complication that the structural producers of the harms are now technical artefacts whose maintenance is distributed across institutional, commercial, and governmental actors.

BIBLIOGRAPHY

- Bender, S. (2023). Diversity and language technology: How language modeling bias causes epistemic injustice. *Ethics and Information Technology*, 25(1), 1–13. <https://doi.org/10.1007/s10676-023-09742-6>
- Chowdhury, S. (2024). Technology is never neutral: Robodebt and a human rights analysis of automated decision-making on welfare recipients. *Australian Journal of Human Rights*, 30(1), 20–40. <https://doi.org/10.1080/1323238X.2024.2409620>
- Cummings, S., Munthali, N., & Shapland, P. (2023). Doing epistemic justice in sustainable development: Applying the philosophical concept of epistemic injustice to the real world. *Sustainable Development*, 31(4), 1965–1977. <https://doi.org/10.1002/sd.2497>
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
- Kerfoot, C., & Bello-Nonjengele, B. O. (2023). Towards epistemic justice: Constructing knowers in multilingual classrooms. *Applied Linguistics*, 44(3), 462–484. <https://doi.org/10.1093/applin/amac049>
- Milano, S., & Prunkl, C. (2025). Algorithmic profiling as a source of hermeneutical injustice. *Philosophical Studies*, 182(1), 185–203. <https://doi.org/10.1007/s11098-023-02095-2>
- Nikidehaghani, M., Andrew, J., & Cortese, C. (2023). Algorithmic accountability: Robodebt and the making of welfare cheats. *Accounting, Auditing & Accountability Journal*, 36(2), 677–711. <https://doi.org/10.1108/AAAJ-02-2022-5666>
- Smith, H. (2020). Algorithmic bias: Should students pay the price? *AI & Society*, 35(4), 1077–1078. <https://doi.org/10.1007/s00146-020-01054-3>
- Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society*, 9(2), 1–16. <https://doi.org/10.1177/20539517221115189>
- Symons, J., & Alvarado, R. (2022). Epistemic injustice and data science technologies. *Synthese*, 200(2), 87. <https://doi.org/10.1007/s11229-022-03631-z>
- van Bekkum, M., & Borgesius, F. Z. (2021). Digital welfare fraud detection and the Dutch SyRI judgment. *European Journal of Social Security*, 23(4), 323–340. <https://doi.org/10.1177/13882627211031257>
- van Toorn, G., & Scully, J. L. (2024). Unveiling algorithmic power: Exploring the impact of automated systems on disabled people's engagement with social services. *Disability & Society*, 39(11), 3004–3029. <https://doi.org/10.1080/09687599.2023.2233684>

EPISTEMIČKA NEPRAVDA U ALGORITAMSKOM ODLUČIVANJU U SISTEMIMA SOCIJALNE ZAŠTITE: OPERACIONALIZACIJA FRICKERINE TESTIMONIJSKE I HERMENEUTIČKE NEPRAVDE ZA AUTOMATIZOVANO CASE-MANAGEMENT

Reda Mansouri

Univerzitet Alžir 2 „Abou El Kacem Saâdallah”
Alžir, Narodna Demokratska Republika Alžir
E-mail: reda.mansouri@univ.alger2.dz

Primljeno: 02.08.2025. Prihvaćeno: 17.10.2025.

Originalni naučni članak

DOI: <https://doi.org/10.65932/CR-2025-2-5> UDK: 1 Fricker M.: [165:17]

Sažetak: Članak se bavi preciznim pitanjem koje se javlja na sjecištu teorije epistemičke nepravde Mirande Fricker i sistemske automatizacije socijalno-zaštitnog case-management procesa u Evropi, Velikoj Britaniji i Australiji: gubi li korisnik svoju epistemičku poziciju kada je kategorizovan algoritmom čiji interpretativni horizont ne dijeli i čiji kategorijalni aparat ne može pregledati? Oslanjajući se na temeljnu Fricker distinkciju između testimonijalne nepravde (deficit kredibiliteta dodjeljen govorniku na osnovu identitetske predrasude) i hermeneutičke nepravde (strukturni deficit u zajedničkim interpretativnim resursima koji marginalnu grupu stavlja u nepovoljan položaj pri artikulaciji svog iskustva), rad razvija tipologiju pet modova algoritamskog epistemičkog zatvaranja kojima automatizovani case-management sistemi sistematski reprodukuju oba oblika nepravde. Pet modova su: (1) kategorijalna translacija, kojom se korisnikova narativa kompresuje u pre-definisana sistemska polja koja prihvataju samo one vrste svjedočenja koje algoritam može prepoznati; (2) kalibracijsko utišavanje, kojim algoritam istreniran na pristrasnim historijskim podacima sistematski diskontuje određene klase svjedočenja kao statistički manje kredibilne; (3) opacitet (neprozirnost), kojim black-box karakter modela uskraćuje korisniku interpretativne resurse potrebne za razumijevanje odluke; (4) vremenska kompresija, kojom brzina automatizovane obrade onemogućava prostor za nijansiranje, kontekstualizaciju i follow-up; i (5) asimetrija žalbe, kojom formalne žalbene procedure pretpostavljaju razinu algoritamske pismenosti i infrastrukturnog pristupa koji pogođeni korisnik karakteristično nema. Tipologija se empirijski primjenjuje na tri dokumentovana slučaja — holandski SyRI sistem za detekciju zloupotrebe socijalne pomoći, britanski Ofqual algoritam za ocjenjivanje A-level ispita iz 2020. godine, i australski Robodebt sistem za automatsku naplatu duga — oslanjajući se na recenzirane analize tih slučajeva (van Bekkum i Borgesius, 2021; Smith, 2020; Nikidehaghani, Andrew i Cortese, 2023). Članak argumentuje da pet modova ne djeluju izolovano; oni formiraju integrisan aparat čiji kumulativni efekat je konverzija osobe koja traži socijalnu zaštitu u ono što ja zovem “algoritamski nečitljiv subjekat” — pozicija u kojoj se ni svjedočenje ni samo-interpretacija ne mogu uspješno predočiti unutar sistema u njegovom postojećem obliku.

Ključne riječi: *epistemička nepravda, Miranda Fricker, testimonijalna nepravda, hermeneutička nepravda, algoritamsko odlučivanje, socijalna zaštita, automatizacija socijalne pomoći, SyRI, Ofqual, Robodebt, algoritamsko epistemičko zatvaranje, algoritamski nečitljiv subjekat.*