Video-The Evidentiary Burden

By NotebookLM

For nearly four centuries, Harvard University had never revoked the tenure of a sitting professor. That institutional precedent ended with the case of Francesca Gino. This dismissal was executed through a Third Statute proceeding, a rarely used university mechanism designed specifically to strip a faculty member of their position.

To justify such a move, the university must meet a strict legal threshold. It cannot act on simple probability. It must establish misconduct through clear and convincing evidence.

This scale shows how burdens of proof are distributed. A standard civil case requires a preponderance of evidence, just over 50%. The clear and convincing standard is significantly higher, requiring a certainty level of 75% to 80%.

In a case of academic misconduct, this standard requires investigators to prove that a researcher acted with intent. The simple presence of data anomalies is not enough. The fabrication must be shown to be deliberate. If the empirical evidence allows for alternative explanations, such as research noise, coding errors, or administrative oversight, the clear and convincing standard is not legally satisfied.

The process began after the group Data Colada notified the university of their findings, threatening to publish their allegations if the Harvard Business School did not act immediately. In response, the Dean of the business school bypassed the faculty-approved procedures for investigating misconduct, launching a new ad hoc process created specifically for this case.

An audit of this procedure suggests that its methodological framework resulted in misclassifying standard research noise as evidence of intentional fraud. A multimillion-dollar institutional audit is only as reliable as the baseline procedural assumptions it starts with. Determining whether those assumptions were sound is critical to evaluating the final verdict.

The ad hoc procedure introduced a significant imbalance from the start, placing Gino under a strict gag order for 16 months while the university built its case. While Harvard's committee used a paid

forensic data analyst to analyze the research, the university's rules explicitly prevented the defense from hiring its own expert during the investigation.

This timeline shows the asymmetry. The university's track moved forward for over a year, culminating in 180-page forensic report. The defense track remained locked, unable to review the data with experts until after a determination of guilt had been made. When Gino was finally permitted to hire experts, they identified significant errors in the university's analysis. Harvard eventually withdrew its own report, but then introduced a new expert analysis, giving the defense only 30 days to respond.

This process also intersected with the Department of Health and Human Services rule, which establishes a six-year statute of limitations for investigating research misconduct. Three of the four papers in the investigation were over six years old. The university maintained it could investigate them by arguing that simply listing the papers on a website constituted a continued use of the research.[1]

These procedural shifts restricted the defense's ability to utilize the time, expert resources, and temporal protections typically required to reach a factual conclusion.

The hearing committee's argument for guilt centered on a single logical claim that every identified data alteration across the papers strengthened the research hypotheses. This logic is central to the prosecution. If every error benefits the researcher's conclusion, it creates a mathematical inference of intent. However, an empirical analysis of the source data shows this claim to be inaccurate.

Take the data for the 2020 networking paper, which contained 1066 anomalous entries flagged by investigators. Only 62% of these changes actually support the hypothesis. The remaining 38%,[2] 415 entries, have no impact on the paper's conclusion.

In the Allegation 2 dataset, the number of irrelevant changes is even higher, reaching 44%. A researcher intentionally optimizing a dataset generally does not manually alter hundreds of cells that have no effect on the outcome. This mixed pattern is more consistent with procedural or software errors than with targeted manipulation.

---

[1] Clarification: The university argued that self-citations counted as renewed use, even if the citations were generic citations in a broad literature review.
[2] Correction: it is 61% and 39%, not 62% and 38%. 39% of the changes affected variables irrelevant to the study hypothesis.

In Allegation 4, investigators focused on the study conducted 16 years ago at the University of North Carolina. The investigators compared two spreadsheets, File A, which they assumed was the raw data, and File B, which they claimed contained fraudulently added participant rows. The basis for this claim changed when physical paper receipts for participant payments were discovered in archival storage.[3] Cross-referencing the paper receipts showed that File A was not raw data at all. It was an incomplete workspace file, missing a large number of the actual study participants.[4] The final analytical file, File B, matched the physical paper receipts perfectly.[5] The rows were not added; they were simply missing from the incomplete file the university used as its baseline. Proof of fraud requires an accurate comparison by selecting an incomplete midpoint file as the baseline. The investigation generated a false positive for data manipulation.

The investigation also overlooked the operational pipeline of modern empirical research. Gino employed over 60 research assistants who handled, cleaned, and coded data before any final statistical analysis took place.

This chart tracks the data handlers in the pipeline. Data moves from the software export through multiple assistants before reaching its final state. Despite their role in cleaning the data, Harvard investigators only interviewed two of the 12 research assistants directly involved in these specific papers.

This omission was critical in Allegation 3. Investigators found differences between a column for lying and a column for cheating and assumed they were fraudulent changes made by the principal investigator. Because they did not interview the assistants, they failed to establish if the RAs used broader criteria to define cheating, such as looking up answers online, even if a participant hadn't lied about a coin flip.

This lack of pipeline analysis also impacted Allegation 2, where investigators initially claimed 20 rows of data were fabricated by Gino. The university eventually conceded that those 20 rows were actually entries from automated bots, a fact missed by the investigation's original analysis.

---

[3] Clarification: they were discovered in boxes with materials from Gino's HBS office.

[4] Correction: a few participants, not a large number of them.

[5] Clarification: every participant in File B has a match in the receipts, while not all participants in File A have a match in the receipts.

When a process misidentifies bot noise and standard data cleaning as intentional fraud, the accuracy of the overall audit is undermined. The investigation into Francesca Gino reached a finding of fraud by bypassing basic fact finding, including research assistant interviews and overlooking exculpatory paper records.

Returning to the legal standard, a third statute removal requires 75% to 80% certainty of intentional misconduct. This certainty must be weighed against the empirical record, the mixed pattern of anomalies, the reliance on incomplete baseline files, and the misidentification of automated bot traffic. Factoring in these methodological errors causes the level of certainty to shift. It falls back toward a simple preponderance, well below the mandatory threshold for revocation. Large data pipelines will always contain noise, but the clear and convincing standard exists specifically to distinguish between these structural errors and intentional manipulation.

The evidence presented in this proceeding does not meet the strict legal standard required to end a tenured professor's career. The case remains a significant study in the tension between institutional processes and the rigorous demands of science.