Management as Rules: Evidence from an adaptive Bayesian questionnaire *

Tom Schwantje

Abstract

Organisational practices are often communicated and implemented through rules, stable "if–then" decision policies, yet such rules have been difficult to measure at scale. This paper develops a novel adaptive Bayesian questionnaire to measure managers' use of rules within organisations. The survey first elicits decisions across a set of hypothetical scenarios, and then elicits the rules managers report using based on patterns in those decisions. This yields a firm-level measure of the share of decisions governed by formal rules and the content of those rules. I apply this tool to senior HR managers at 284 medium-sized and large Ethiopian firms. I find that the use of formal rules to make decisions is positively and robustly associated with firm performance and explains variation in performance across firms above and beyond a standard measure of management practices. Using firms' exposure to COVID-19 disruptions, I provide suggestive evidence of a trade-off: rule-based management is associated with stronger performance in stable periods but weaker performance during shocks. These findings motivate further empirical research on the causal mechanisms linking organisational rules to performance and why organisations choose to adopt rules.

Keywords:
Firms, Management, Rules, Bayesian Survey Design, Development

*I am grateful to Simon Quinn for his support. I want to thank Girum Abebe, Caitlin Brown, Dennis Egger, Douglas Gollin, Stephen Hansen, Alison Jackson, Michael Koelle, Julien Labonne, Rocco Macchiavello, Muhammad Meki, Christian Meyer, Christopher Woodruff. I also thank seminar audiences at the University of Oxford, the 2nd Workshop on the Frontiers in Measurement and Survey Methods and the Empirical Management Conference at Stanford. I also want to thank Gezahegn Gbremedhin for outstanding research assistance. The project would not have been possible without the constant support of Rose Page and the Centre for the Study of African Economies (University of Oxford), nor without the support of EconInsights in Addis Ababa. IRB approval was obtained from the University of Oxford (ECONCIA19-20-13.)

1 Introduction

Large and persistent productivity gaps across firms – even within narrowly defined industries – remain a central puzzle in economics (Syverson, 2011). While existing work shows that better management practices explain part of this variation (Bloom and Van Reenen, 2007; Gibbons and Henderson, 2012), we lack systematic evidence on the microfoundations of these organisational practices, many of which are operationalised through rules and procedures. For example, widely adopted management philosophies such as scientific management and lean management (Press, 1989; Taylor, 1911), are implemented through (often codified) systems of rules – yet we lack empirical evidence on whether differences in such rules help explain why otherwise similar firms perform so differently.

Economic and organizational theory suggests two channels through which rules affect firm performance. First, rules reduce the costs of delegation: by specifying how decisions should be made, they allow senior managers to extend their span of control without sacrificing consistency (Aghion and Tirole, 1997; Bloom, Genakos, Sadun, and Van Reenen, 2012). Second, rules support coordination: shared decision procedures help employees act coherently without constant communication (Gibbons and Henderson, 2012; March and Simon, 1958). Yet the same rigidity that enables coordination may limit adaptation. When circumstances change, rules that once enhanced efficiency can become constraints – preventing managers from responding to local information or novel situations (Adler and Borys, 1996; Aghion, Bloom, Lucking, Sadun, and Van Reenen, 2021; Li, Mukherjee, and Vasconcelos, 2022). This implies a fundamental trade-off: rules enhance performance in stable environments but may reduce resilience when firms face shocks (Englmaier, Galdon-Sanchez, Gil, Kaiser, and Strandt, 2025). Testing these predictions empirically, however, has proven difficult.

Organisational rules – shared "if–then" procedures for recurring decisions – have received limited empirical attention for three reasons. First, rules are hard to measure empirically as they are fundamentally high-dimensional: the number of potential rules that might govern any particular factual scenario is very large, and increases exponentially with the number of facts that a decision maker might consider. Second, a rule is defined as much by the information it *does* use, as by the information it *does not*: for example a rule may dictate an applicant's gender *cannot* be considered when making a decision. Third, observing realised decisions alone does not allow one to identify whether a decision pattern reflects a shared organisational constraint or individual managerial discretion, nor does it reveal which information the manager treated as relevant. These features imply that measuring rules requires recovering both decision mappings and the information managers treat as admissible, and doing so in a way that distinguishes shared procedures from individual discretion.

In this paper, I develop a novel empirical tool to measure such organisational rules based on a vignette-based elicitation method. I implement this approach to survey 284 senior managers at medium-sized and large Ethiopian firms and construct a measure of rule-based management in these firms. I find that firms that rely more on formal (documented) rules to make decisions have a higher profit margin, which is robust to controlling for a large set of firm observables. Moreover, I show that this measure is more predictive of firms' margin ratio, and profitability, than a management score based on the Management and Organisational Practices Survey (MOPS) Bloom, Brynjolfsson, Foster, Jarmin, Patnaik, Saporta-Eksten, and Van Reenen (2019). Finally, leveraging firms' exposure to COVID-19 disruptions, I provide suggestive evidence that firms with more rule-based management were more negatively affected by this economic shock—consistent with a trade-off in which rules enhance efficiency under stable conditions but reduce flexibility when circumstances change.

First, I develop a methodology to overcome the three challenges identified in measuring rules. The central idea is straightforward: if a manager relies on rules, her decisions should follow consistent patterns: specific combinations of facts lead to specific actions. To elicit these patterns, I present respondents with a small number of carefully selected hypothetical decision problems and record their choices. Each vignette presents a decision – such as whether to promote an employee – and varies key facts like performance or tenure to learn about patterns in a respondent's choices. I also record which information respondents use to make a decision, revealing both what enters their decision and what information they treat as irrelevant. Finally, I confirm that these patterns reflect organisational rules rather than individual discretion by asking whether they are driven by the respondent's own preferences or by decision procedures commonly used within the organisation. This approach allows me to measure rule-based behaviour without needing managers to describe those rules explicitly.

A key feature of this methodology is that the sequence of decision problems is chosen adaptively: each question is chosen to be as informative as possible about the respondent's decisions. This adaptive design addresses the dimensionality challenge: rather than attempting to cover an exponentially large space of possible decision problems, the survey concentrates questions on situations that most help explain patterns in a respondent's decisions. To do so, the survey updates beliefs about how the respondent is likely to decide in other cases and then chooses the next case that is expected to reduce uncertainty the most. This adaptive design builds on the Bayesian optimisation literature (Frazier, 2018) and, as I show in Section 7, substantially increases the rate of learning relative to a fixed questionnaire.

I implement this method in a survey of senior HR managers at medium-sized and large Ethiopian firms. As the sample spans a wide range of industries and firm sizes, I focus on HR managers. This allows me to study relatively comparable decisions – related to compensation and hiring – across this heterogeneous sample of firms. Organisational rules are likely particularly relevant for firms in low-income countries, which face high costs of supervision and delegation (Bassi, Lee, Peter, Porzio, Sen, and Tugume, 2023), and in particular in Ethiopia where we find a rule-based management style is highly valued among young professionals (Abebe, Fafchamps, Koelle, Quinn, and Schwantje, 2025).

Turning to results, I first document basic properties of the rules I recover. The rules appear sensible: in the pay-rise scenario, for instance, the facts most commonly entering rules are employee performance and tenure. They are also parsimonious: firms typically rely on only a small subset of available information. Finally, I distinguish formal (documented) from informal (commonly known but undocumented) rules and show that firms differ markedly in which type they rely on. For each firm, I define their rulebook as the set of rules they use in each scenario.

As primary measure of the extent of rule-based management firms implement, I define the breadth of the rulebook. The breadth of the rulebook is defined as the share of decisions in a specific scenario – for example, the pay rise scenario – for which the manager says that they use a rule to make a decision. I show that in particular better managed firms, and plants which are a subsidiary of a larger firm, use a more rule-based style of management.

Next, I consider the (correlational) link between the adoption of a rule-based style of management and firm performance. I show that firms which have a broader rulebook report higher profits, and in particular higher profit margins suggesting greater operational efficiency. To test whether a more rule-based style of management is simply correlated to other factors that imply the firm is more productive, I show that this relationship is robust to controlling for a rich set of firm characteristics. Notably, controlling for the breadth of the firms' formal rulebook has clear incremental predictive power for the profit ratio above and beyond standard management indices such as the MOPS score (Bloom et al., 2019), suggesting it captures something separate – and important – about how firms are organised.

Finally, I test theoretical prediction that adopting a more rule-based management style increases performance at the cost of decreased resilience. To provide suggestive evidence, I show that firms with a more rule-based style of management were more negatively affected by COVID-19 restrictions in Ethiopia. This pattern is consistent with theoretical work suggesting that rigid decision structures limit adaptability under uncertainty (Aghion et al., 2021; Li et al., 2022). Together, the results provide additional empirical support for the prediction that rules may support performance in stable environments but constrain flexibility when conditions change.

I make three contributions to the literature. First, I develop and apply a Bayesian adaptive questionnaire to feasibly measure organisational rulebooks—that is, the codified decision rules and approval thresholds that managers apply to recurring choices. This advances the literature that seeks to open the black box of management, including work on management practices (Bloom and Van Reenen, 2007), relational contracts (Gibbons and Henderson, 2012), CEO behaviour (Bandiera, Prat, Hansen, and Sadun, 2020), autonomy (Aghion et al., 2021), and delegation (Akcigit, Alp, and Peters, 2021a). The ability to measure organisational rules opens a new empirical research agenda in management and relational contracting. By providing a tool to quantify rules as micro-level mechanisms of formalization, it offers microfoundations for theories of when rules enable coordination versus constrain adaptation (Adler and Borys, 1996; Aghion and Tirole, 1997). The developed methodology also allows researchers to study questions related to better designing and targeting productivity-enhancing interventions, such as management practices or access to capital.

Second, I contribute to the literature on management practices in firms in low-income countries. I show that the implementation of rules is a central component of managerial activity (Gibbons and Henderson, 2012), and that measuring rules helps explain heterogeneity in firm performance beyond standard management indices (Atkin, Khandelwal, and Osman, 2019; Syverson, 2011). This is particularly important in low-income countries, where the cost of delegation is typically very high (Akcigit, Alp, and Peters, 2021b; Bloom, Eifert, Mahajan, McKenzie, and Roberts, 2013; Bloom et al., 2012)—to the extent that rules reduce these costs. This aligns with the suggestion of Bassi et al. (2023) that reducing the cost of delegation and increasing managers' span of control are crucial to unlocking economies of scale in sub-Saharan Africa. Using firms' exposure to COVID-19 disruptions, I corroborate the finding of Englmaier et al. (2025) that more rigid, rule-based management practices entail a trade-off between performance and resilience—such practices perform well in stable periods but underperform during economic shocks. More broadly, by measuring the decision logics managers actually apply, this paper provides microfoundations for classic organisational design arguments about when formalization enables coordination versus when it constrains adaptation (Adler and Borys, 1996; Aghion and Tirole, 1997).

Third, the Bayesian adaptive questionnaire I develop differs from existing methods in a fundamental way by focusing on maximizing learning about a set of related decision outcomes, rather than identifying specific parameters or locating a single optimum. Existing methodologies either focus on finding a maximum (Frazier, 2018; Li, Raymond, and Bergman, 2020), identifying specific parameters (Chapman, Snowberg, Wang, and Camerer, 2024), estimating treatment effects (Caria, Gordon, Kasy, Quinn, Shami, and Teytelboym, 2024; Kasy and Sautmann, 2021), or learning about a one-dimensional function (Callander and Matouschek, 2019). Methodologically, the innovation in this paper is to apply principles from Bayesian optimisation to learn about a set of related decision problems. This method is more flexible than existing approaches to optimal learning and performs comparably to recent methods in adaptive choice experiments.

The remainder of the paper proceeds as follows. Section 2 provides details on the Bayesian adaptive experimentation algorithm and on how this algorithm is implemented to measure organisational rules. Section 3 describes the sample and the field implementation of the tool. Section 4 examines how rules differ across firms and managers. Sections 5 and 6 explore the relationship between rules and firm performance and resilience to economic shocks, respectively. Section 7 evaluates the performance of the adaptive sampling algorithm using both simulated and survey data. Finally, Section 8 concludes.

2 Measuring Organisational Rules

To measure organisational rules empirically, I exploit a simple insight: if a manager follows rules when making decisions, their choices should exhibit systematic patterns—specific combinations of facts trigger specific actions, and some facts are treated as irrelevant and therefore ignored. I operationalise this in three steps. First, I present managers with a sequence of hypothetical decision problems (cases). For each case, I record both their decision and which attributes they choose to observe before deciding, revealing not just what they do but what information they treat as decision-relevant. Second, I reverse-engineer candidate rules from these patterns as Boolean "if–then" statements – patterns linking combinations of observable facts to actions that could be driven by a single rule (e.g., "if an employee is high-performing and low-paid, grant a pay rise"). Third, I ask managers to validate each candidate rule: first, whether it reflects a rule used in the organisation; and if so, whether it is a formal documented policy, an informal but commonly known practice, or a rule they apply personally. Only rules confirmed as formal or informal organisational practice enter my measure of rule-based management.

This approach addresses three measurement challenges. First, the space of possible rules grows exponentially with the number of case attributes: with d binary characteristics there are 2^d possible cases and 2^{2^d} possible Boolean rules. I address this using adaptive scenario selection that concentrates questions on informative cases (Section 2.3). Second, rules are defined as much by the information they exclude as by what they use; allowing respondents to choose which attributes to observe directly reveals perceived relevance (Section 2.4). Third, consistent choice patterns may reflect individual preferences rather than shared organisational constraints; explicit validation distinguishes rules from discretion (Section 2.6). The next subsection formalises the definition of rules and rulebooks; subsequent subsections detail each methodological step.

2.1 Defining Rules and Rulebooks

I define a decision rule as a Boolean function that maps a subset of observable case attributes to a binary decision. Let $x = (x_1, \ldots, x_D) \in \{0,1\}^D$ denote a case described by D binary characteristics, and let $y \in \{0,1\}$ denote a decision. A rule $h : \mathcal{D}_h \subseteq \{0,1\}^D \to \{0,1\}$ specifies a decision for every case in its domain $\mathcal{D}_h$. For instance:

> *"If an employee is high-performing, low-paid, and hard to replace, grant a pay rise."*

This rule applies to all cases meeting these three conditions, regardless of other attributes such as tenure or gender.

A defining feature of rules is that they depend on only a subset of available information. Formally, a rule *depends only on* attributes $S_h \subseteq \{1, \ldots, D\}$ if $h(x) = h(x')$ whenever $x_i = x'_i$ for all $i \in S_h$—cases that agree on the relevant attributes yield the same decision regardless of other characteristics. This feature is substantively important: a rule that excludes gender from consideration enforces a constraint that observed decisions alone cannot reveal.

A *rulebook* $\mathcal{R} = \{h_1, \ldots, h_k\}$ is the collection of both formal and informal rules used by the organisation. I measure the *breadth of the rulebook* as the share of possible cases for which at least one rule applies:[1]

$$B(\mathcal{R}) = \frac{|\mathcal{D}_\mathcal{R}|}{2^D}, \quad \text{where} \quad \mathcal{D}_\mathcal{R} = \bigcup_{j=1}^{k} \mathcal{D}_{h_j}.$$

Breadth captures the extent to which decisions are governed by shared organisational rules rather than left to individual discretion. A breadth of zero means no cases are rule-governed; a breadth of one means rules fully determine decisions across all possible cases.

2.2 From Decisions to Rules

If a manager implements a rulebook $\mathcal{R}$, her decisions for all cases $x \in \mathcal{D}_\mathcal{R}$ follow from these rules. In principle, we could ask about every possible case, extract patterns, and validate which reflect organisational rules. I illustrate this logic with a simple example before turning to the challenge that motivates adaptive sampling.

Consider the pay-rise scenario introduced above, with three binary attributes: $x_1 = 1$ indicates high performance, $x_2 = 1$ indicates low pay, and $x_3 = 1$ indicates high mobility (the worker could easily leave). With $d = 3$ attributes, there are $2^3 = 8$ possible cases. Suppose we ask a manager about all eight and observe that she grants a raise whenever performance is high, and never when performance is low. This pattern is consistent with several candidate rules—for instance, "always grant a raise when $x_1 = 1$" or "never grant a raise when $x_1 = 0$." To determine whether these reflect organisational rules or personal discretion, I validate each candidate with the respondent.

The problem is scale. With $d = 7$ attributes there are $2^7 = 128$ possible cases—too many to ask exhaustively. More generally, the number of possible Boolean rules is 2^{2^d}, which grows rapidly.

[1] Similarly, the breadth of a single rule is the share cases to which a single rule applies.

This motivates an adaptive approach: rather than asking about every case, I select each question to be maximally informative about the manager's decision function, and allow respondents to reveal which attributes they treat as relevant.

2.3 Adaptive Scenario Selection

To address the dimensionality challenge, I develop a Bayesian adaptive questionnaire that selects each case to be maximally informative about the respondent's decisions. The intuition is that similar cases elicit similar responses: once we observe a manager's answer to one case, we can update beliefs about how she would decide in related cases. The algorithm exploits this structure by targeting cases where uncertainty is highest.

Formally, let $C_{t-1} = \{(x_i, y_i)\}_{i=1}^{t-1}$ denote the cases observed through question $t-1$. The objective is to select case x_t to minimise expected posterior entropy over all 2^d cases after observing the response:

$$x_t^* = \arg\min_{x \in X} \mathbb{E}_{y \sim P(\cdot|x, C_{t-1})} \left[\sum_{x' \in X} H(y_{x'} \mid C_{t-1} \cup \{(x, y)\}) \right], \tag{1}$$

where $H(y_{x'} \mid C) = -\sum_{y \in \{0,1\}} P(y \mid x', C) \log P(y \mid x', C)$ is the entropy of the predicted decision at case x'. Intuitively, we ask the question whose answer most reduces uncertainty about how the respondent would decide across all possible cases.

This objective is well-suited for identifying decision rules. Cases with high predictive entropy are precisely those where plausible decision rules disagree: if competing rules yield different predictions for case x, the model is uncertain about y_x. By targeting high-entropy cases, the algorithm efficiently discriminates among candidate rules.

To implement this objective, I approximate the unknown decision function using a Gaussian Process (GP) classifier. The GP provides a flexible surrogate that generalises from sparse observations while quantifying uncertainty over unobserved cases. A logistic link maps the latent function to choice probabilities: $P(y = 1 \mid x) = \sigma(f(x))$. Posterior updating uses the Laplace approximation; technical details on kernel specification are in Appendix D. With $D = 7$ attributes, the case space ($2^7 = 128$ cases) is small enough to evaluate the objective exactly by enumeration.

The algorithm proceeds as follows: start with an initial case, observe the manager's decision, update the GP posterior, then select the next case minimising expected entropy. This continues for a fixed number of questions (eight in the implementation).

A key feature is that respondents choose which attributes to observe (Section 2.4), so each response provides information about multiple cases. When a respondent observes only attributes $S \subset \{1, \ldots, D\}$ and makes a decision, that decision is assumed to apply to all $2^{D-|S|}$ cases sharing the same values on S—the unobserved attributes are treated as irrelevant to that choice. The GP conditions on this partial input and extrapolates accordingly, allowing the model to learn both what decisions are made and which attributes are decision-relevant.

2.4 Revealing Information Use

An additional feature of this methodology is that respondents choose which attributes to observe before deciding. For each case, the respondent sees no attribute values initially and requests only those she considers relevant. This design serves two purposes. First, it mirrors real-world decision-making, where managers prioritise a subset of available information. Second, it directly addresses the second key measurement challenge: rules are defined as much by the information they exclude as by what they use. If a manager consistently decides without observing an attribute—such as gender or connections—this reveals that the attribute is inadmissible under her decision procedure, something observed decisions alone cannot show.

Formally, let $S \subseteq \{1, \ldots, D\}$ denote the indices of attributes the respondent chooses to observe for a given case. The decision y is interpreted as a function of only these observed attributes. The GP model conditions on the partial input $x_S = (x_i)_{i \in S}$ when updating posterior beliefs; unobserved attributes are treated as uninformative for that response. To reduce cognitive burden, respondents may specify default attributes they always wish to observe, with the option to request additional information as needed.

This design ensures the methodology captures not only what managers decide, but what information they treat as decision-relevant—a direct window into the structure of their decision rules.

2.5 Extracting Candidate Rules

From the observed decisions, I extract candidate rules—Boolean conditions consistent with every response. The key insight is that each decision covers not just the single case presented, but all cases sharing the same values on the attributes the respondent chose to observe. If a manager observes only performance, pay, and mobility (3 of 7 attributes) and grants a raise, this single response covers $2^{7-3} = 16$ cases—all combinations of the four unobserved attributes yield the same decision.

To illustrate, suppose a manager makes the following two decisions in the pay-rise scenario:

- Observes $(x_1 = 1, x_2 = 1, x_3 = 1)$—high-performing, low-paid, mobile—and grants a raise.
- Observes only $(x_1 = 0)$—not high-performing—and denies a raise.

The first decision covers 1 case (all attributes observed). The second covers 2^{d-1} cases (all combinations of pay and mobility with $x_1 = 0$). Together, these two responses may cover a substantial share of the case space.

A candidate rule is any Boolean "if-then" statement whose prescribed action matches the manager's decision for every covered case in its domain. From the example above, valid candidates include "if high-performing, low-paid, and mobile, grant a raise" and "if not high-performing, deny a raise." Many candidate rules typically satisfy this consistency requirement, including narrower rules nested within broader ones.

I rank candidates from broadest (fewest attributes, most cases covered) to narrowest. Broader rules represent stronger organisational constraints—they remove more discretion. Up to ten of the broadest candidates are presented for validation; once a broad rule is confirmed, narrower nested rules are skipped.

2.6 Validating Candidate Rules

This step addresses the third measurement challenge: consistent decision patterns may reflect individual preferences rather than shared organisational constraints. Asking managers directly about rules risks experimenter-demand effects; eliciting decisions alone cannot distinguish rules from discretion. By combining both—requiring consistency in decisions before validating rules—I discipline what managers can claim as organisational practice.

For each of the ten broadest candidate rules, the manager indicates: (a) whether it reflects a rule used in her organisation, and if so, (b) whether it is a formal documented policy or an informal but commonly known practice. Once a broad rule is confirmed, narrower rules nested within it are skipped—they would not add to the rulebook's breadth. Only rules explicitly affirmed as organisational practice enter my measure; patterns attributed to personal discretion are excluded.

I cap validation at ten rules to avoid respondent fatigue. This design choice has a consequence: when a manager follows many narrow rules leading to the same action, they merge into one broader rule, so the measure captures the breadth of the rulebook rather than the number of distinct clauses. Conversely, a manager applying many narrow rules with different actions across similar cases will yield a rulebook with narrow measured breadth. This is consistent with my conceptualisation of rules as stable, commonly understood guidelines rather than case-by-case protocols.

2.7 Algorithm Performance

To validate the adaptive sampling algorithm, I conduct three analyses: (1) comparing performance against random sampling using pilot data, (2) simulations across different data-generating processes, and (3) benchmarking against the DOSE algorithm from Chapman et al. (2024). The adaptive algorithm increases learning efficiency by 50–300% relative to random sampling. For readers interested in this methodological validation, the full results are in Section 7.

3 Management as rules in Ethiopian firms

I implement this methodology to measure the use of rules by HR managers at Ethiopian firms. To do so I use two scenarios, one related to compensation and one to hiring. The key question for the respondent is whether to respectively give a pay rise or hire an employee as a function of seven binary characteristics. This section first details the exact scenarios and provides a further example, before turning to the exact sample used in the survey and the details of the implementation.

3.1 The Human Resource Scenarios

Box 1 and Box 2 detail the exact pay rise and hiring scenario respectively. In these scenarios, a colleague – role-played by the enumerator – asks for the respondent's advice. This approach creates a realistic context in which the respondent knows they cannot consider all of the employee's characteristics when making a decision. In both scenarios, we frame the discussion by asking the respondents to identify the most common position in their firm, and to answer the questions as if relating to this position.

The pay rise scenario (Box 1) is presented as follows: "Suppose I, as your colleague, tell you an employee is unhappy with their pay and would like a rise to the next point on the pay scale, and I ask for your decision. Before making a decision, you can ask to learn about any of the following characteristics of the employee." The enumerator then reads out the seven characteristics in a random order, and shares a document with the respondent listing the full set of scenario characteristics for reference. The specific characteristics, for both scenarios, were selected based on piloting the module in Ethiopia.

This box also provides an example of how the survey may play out in practice. The column "example case" provides a draw of seven binary characteristics that may be generated by the adaptive algorithm, and given to the respondent. This example case fixed the seven realisations of the individual characteristics. The respondent can then, iteratively, choose which of these individual characteristics to observe. For example, as indicated in the observed column the respondent may wish to ask about performance, mobility and pay, observe the fixed characteristics for this case, and then make a decision to not give a pay rise. If this were the only question asked, this would yield the candidate rule: "If an employee has above average performance but also has above average pay and cannot easily leave for another firm, do not give them a pay rise".

Box 1: The pay rise scenario

"Suppose I, as your colleague, tell you an employee is unhappy with their pay and would like a raise to the next point in the pay scale and I want your decision. Before making a decision, you can ask to learn about any of the following characteristics of the employee."

Abbreviation	Characteristic	Options	Example Case	Observed
Performance	The employee's performance compared to other employees with similar jobs.	Above/below average	Above average	✓
Likes	Your colleague personally likes the employee.	Yes/No	Yes	
Mobility	The employee could easily leave for another firm.	Yes/No	No	✓
Pay Rise	The employee has had a pay rise in the past year.	Yes/No	Yes	
Pay	The employee's pay is above average at the firm.	Yes/No	Yes	✓
Tenure	The employee has worked at the firm for more than two years.	Yes/No	Yes	
Gender	The employee's gender.	Male/Female	Male	

Notes: The pay rise scenario asks the respondent about compensation. The scenario has seven specific characteristics, and the respondent knows what these are. Each characteristic can take two values. The example case is an example case that could be selected by the algorithm. The observed column provides a hypothetical example of which of those fixed characteristics a respondent may choose to observe.

The hiring scenario (Box 2) is presented as follows: "Suppose I, as your colleague, tell you I have a candidate for an entry-level position at your firm for which you are hiring. I would like your decision on whether you would hire this candidate. Before making a decision, you can ask to learn any of the following characteristics." The enumerator then again shares the list of characteristics and reads these out.

Box 2: The hiring scenario

"Suppose I, as your colleague, tell you I have a candidate for an entry-level position at your firm for which you are hiring. I would like your decision on whether you would hire this candidate. Before making a decision, you can ask to learn any of the following characteristics:"

Abbreviation	Characteristic	Options
Enthusiasm	The candidate showed enthusiasm for the job.	Above/below average
Interview	The candidate performed well in the interview.	Yes/no
English	The candidate speaks English.	Yes/no
Culture	The candidate fits with your firm's culture.	Yes/no
Experience	The candidate has relevant experience at another firm.	Yes/no
University	The candidate has a university degree.	Yes/no
Gender	The candidate's gender.	Male/Female

Notes: The hiring scenario asks the respondent about hiring a new employee. The scenario has seven specific characteristics, and the respondent knows what these are. Each characteristic can take two values listed in the "Options" column.

During the survey, the Bayesian adaptive questionnaire selects a sequence of eight cases (a combination of the seven characteristics) to present to the respondent for each vignette. For each case, I observe the respondent's decision and the information they used to make that decision. The algorithm interprets the respondent's decision as indicative of their choice in all cases sharing the observed characteristics. For instance, if the respondent decides to grant a pay rise to an employee with above-average performance and pay but below-average mobility, the algorithm assumes this decision applies to all cases with these characteristics. These cases are added to the set C_o and are used to update the model's beliefs. Based on the accumulated information from previous cases, the algorithm selects the next case to present.

The sequence of decisions made by the respondent and the information used to make these decisions allows the full set of cases to be divided into three groups: the set of cases for which the respondent has indicated they will not give a pay rise, the set of cases for which the respondent has indicated they will give a pay rise, and the remaining cases for which the respondent has not shared a decision. Based on these three sets of cases, I generate a set of hypothetical rules, i.e., mappings from a subset of characteristics to decisions, that are consistent with the respondent's decisions. For example, if an employee performs well, is currently paid relatively poorly, and could easily leave, they should receive a pay rise.

These hypothetical rules form the input for the second part of this questionnaire. In this second part, I directly ask respondents whether these hypothetical rules represent actual rules within their organisation. Starting with the broadest rule, i.e., covering the most cases, I ask the respondent whether it applies. If it does, I next inquire whether the rule is formal (documented and codified) or informal (generally understood but not documented). This process continues with less broad rules as necessary, allowing respondents to classify each of these hypothetical rules.

3.2 The sample

I apply the developed methodology to examine personnel decision rules in a sample of relatively large Ethiopian firms. These firms are representative of businesses in Addis Ababa. Focusing on human resource management–a universal aspect of organisational practices–facilitates the comparison of the diverse set of firms in this sample.

The participating firms are medium-sized to large establishments, operating across various sectors of the economy. The median number of employees is fifty (25th percentile = 20, 75th percentile = 142). Most of the firms are located in Addis Ababa, while around 15% are situated in the mid-sized towns of Adama (80 km from Addis Ababa) and Bishoftu (40 km from Addis Ababa). The firms operate in a range of sectors, primarily manufacturing (40%), services (30%), and retail and trade (20%), with the remainder in other industries. Of the firms sampled, 20% are government-owned, while 80% are privately owned. The sample is drawn from firms previously studied in Abebe, Fafchamps, Koelle, and Quinn (2019), with some additional, comparable firms added to the sampling frame.

The respondents are highly experienced managers responsible for HR decisions within their firms. Around 75 of the surveyed managers are either owners or managing directors, while the rest are senior managers with HR responsibilities. The managers have a median tenure of seven years at their current firms (mean tenure of nine years) and a median of fifteen years of total work experience (mean of eighteen years). Approximately 90% of the managers hold at least a bachelor's degree, and 65% have formal management education. The sample is restricted to managers who have the authority to make decisions related to hiring and compensation within their firm.

3.3 Implementation

A team of experienced enumerators conducted the survey. To ensure clarity and comprehension, the survey focuses on two scenarios, each with seven fixed binary characteristics, resulting in 128 (2^7) possible combinations. For each scenario, respondents were presented with eight distinct cases, each consisting of a unique combination of the seven characteristics. These cases were selected using the adaptive algorithm. The hiring and pay rise scenarios were detailed earlier in Section 3. The respondent is asked to answer the questions as they relate to a position in the main production activity in their firm.

To implement the survey, I developed a custom application to be used by the enumerators in Addis Ababa. Since no reliable internet connection was available, all computations had to be performed locally on the enumerators' computers during the interviews. The application allowed enumerators to record respondents' answers, and then used the Bayesian adaptive algorithm to solve the optimisation problem of selecting the next question. Because the algorithm considers the full history of each respondent's answers – including the information they chose to observe – the sequence of questions was selected in real time. Following this, the survey application generated a set of heuristics consistent with the data collected on decisions. Enumerators then used the application to ask whether these heuristics represented organisational rules. Details of the questionnaire are provided in the Online Appendix B.

Training data for the survey was collected from respondents within the same population by asking about eight cases selected randomly from the full set of potential combinations in the two hypothetical scenarios. These questions were identical to those in the main survey, except that respondents were shown all available information rather than choosing which characteristics to observe. This training data was used to fit the parameters of the Gaussian Process model employed in the survey.

The Gaussian Process model employed a linear mean function and a squared exponential kernel with automatic relevance determination. This configuration allowed the impact of each of the seven characteristics on the covariance between two cases to differ. A logistic link function was used to map the Gaussian Process outputs into probability space.

3.4 Identified patterns in firms' decisions and rules

Turning to the data collected from the adaptive Bayesian questionnaire, I first assess which characteristics respondents chose to observe and how these characteristics were used to define rulebooks. I define a rulebook as the collection of rules identified for a firm.

On average, respondents asked to observe four characteristics per decision (median = 4; 25th percentile = 3, 75th percentile = 5) in both scenarios. The identified rules tended to be broad and typically relied on only two characteristics (median = 2; 25th percentile = 1, 75th percentile = 3). This pattern was consistent across both scenarios. Firms used, on average, two rules per scenario (median = 2; 25th percentile = 1, 75th percentile = 4).

Table 1 summarises the characteristics observed and used in rules for the pay rise scenario. The most frequently observed characteristics were performance and tenure at the firm, while gender was the least observed. A similar trend was found in the probability of each characteristic being included in a rulebook. Interestingly, while performance was often used in decision-making, it featured less frequently in the derived rules. The last two rows in the table consider the probability that each characteristic is included in a rule that leads to a particular decision, as well as the average value of the characteristic when it is part of a rule leading to that decision.[2]

Table 2 presents a summary of the characteristics observed and included in rules for the hiring scenario. In this scenario, a wider range of characteristics were frequently observed, including enthusiasm for the position, interview performance, cultural fit, and experience. Gender was observed more frequently in the hiring scenario compared to the pay rise scenario; however, it was not much more likely to feature in a rulebook.

Table 1: Characteristics used in decision making in the pay rise scenario

	Performance	Like	Mobility	Pay rise	Pay	Tenure	Gender
	Probability a respondent asks about a characteristic in a specific case						
Observed by the respondent	73.7%	46.1%	36.2%	37.7%	33.5%	60.5%	19.1%
	Probability that a characteristic features in a respondents' rulebook.						
Features in a rulebook	54.4%	37.8%	29.5%	20.7%	18.4%	49.8%	14.7%
Decision	Probability of a characteristic featuring in a rule making a specific decision.						
No Raise	40.6%	28.8%	26.5%	22.9%	17.1%	38.2%	12.4%
Raise	31.3%	24.6%	18.7%	11.4%	11.4%	34.5%	10.9%
	Probability of a characteristic being 1 conditional on being in a rule.						
No Raise	24.6%	44.9%	24.4%	43.6%	48.3%	32.3%	47.6%
Raise	92.6%	80.0%	76.4%	72.7%	56.8%	82.0%	61.9%

Notes This table depicts what information managers use in decisions and rules in scenario 1. For the first row, this depicts that probability that a characteristic is observed by a respondent when making a decision. The second row, this depicts the probability that, for each respondent, the characteristic features in at least one rule. The third rows depicts the probability each characteristic features in an individual rule that makes the decision to give the pay rise, and in a rule that makes the decision not to give a pay rise. The final set of results gives the average value of a characteristic, conditional on it featuring in a rule making a specific decision. The characteristics included are performance (1=high), like (1=colleagues likes the respondent), mobility (1 = could leave), pay rise (1 = had one in the last year), pay (1 = above average), tenure (1 = more than two years) and gender (1=male).

[2] Gender is encoded as 0 = female, 1 = male.

Table 2: Characteristics used in decision making in the hiring scenario

	Enthusiasm	Interview	English	Culture	Experience	University	Gender
	Probability a respondent asks about a characteristic in a specific case						
Observed by the respondent	63.6%	64.9%	30.7%	63.9%	74.9%	47.8%	25.5%
	Probability that a characteristic features in a respondents' rulebook.						
Features in a rulebook	46.7%	53.7%	18.5%	46.3%	51.1%	44.1%	15.4%
Decision	Probability of a characteristic featuring in a rule making a specific decision.						
Don't hire	28.9%	26.4%	9.4%	30.2%	32.3%	23.4%	8.1%
Hire	26.1%	29.7%	11.3%	25.9%	29.7%	28.0%	8.8%
	Probability of a characteristic being 1 conditional on being in a rule.						
Don't hire	19.1%	22.6%	36.4%	22.5%	19.7%	36.4%	36.8%
Hire	70.5%	79.1%	37.7%	62.0%	76.3%	77.1%	61.0%

Notes This table depicts what information managers use in decisions and rules in scenario 2. For the first row, this depicts that probability that a characteristic is observed by a respondent when making a decision. The second row, this depicts the probability that, for each respondent, the characteristic features in at least one rule. The third rows depicts the probability each characteristic features in an individual rule that makes the decision to hire the candidate, and in a rule that makes the decision not to hire the candidate. The final set of results gives the average value of a characteristic, conditional on it featuring in a rule making a specific decision. The included characteristics are enthusiasm (1=showed enthusiasm), interview (1=performed well), English (1=speaks English), culture (1=fits with culture), experience (1=has relevant experience), university (1=has a university degree), gender (1=male).

4 The relationship between rules, firms, managers and management

In this section, I examine how rules relate to firm and manager characteristics. I first consider how the use of rules correlates with the characteristics of firms and managers who use those rules. Then, I investigate whether firms with broader rulebooks implement different management practices. To measure the extent to which managers' decisions are constrained by rules, I focus on how comprehensive the rulebook is. Each scenario contains $2^7 = 128$ possible cases. I define the *breadth* of the rulebook as the proportion of cases in which the manager makes their decision using a rule—that is, the number of cases covered by rules divided by the total number of cases in the scenario. A breadth close to one indicates highly rule-based decision-making, whereas a breadth close to zero suggests mainly discretionary decision-making. I calculate the breadth for the entire rulebook and separately for the set of formal rules, defining the *breadth* and the *formal breadth* of the rulebook, respectively.

A rule's breadth is the share of the 2^d scenario space for which its premise is true; a rule that conditions on more attributes is narrower (lower breadth). Stage 1 identifies candidate rules by searching this discrete space; Stage 2 establishes whether each rule stems from firm policy or personal discretion. The adaptive design is essential because managers answer questions about at most eight vignettes, yet with d = 7 there are 128 possible scenarios. Simulations (section 7) show the entropy-based question chooser recovers over 90% of organisational rules while a fixed 8-scenario search would recover fewer than half.

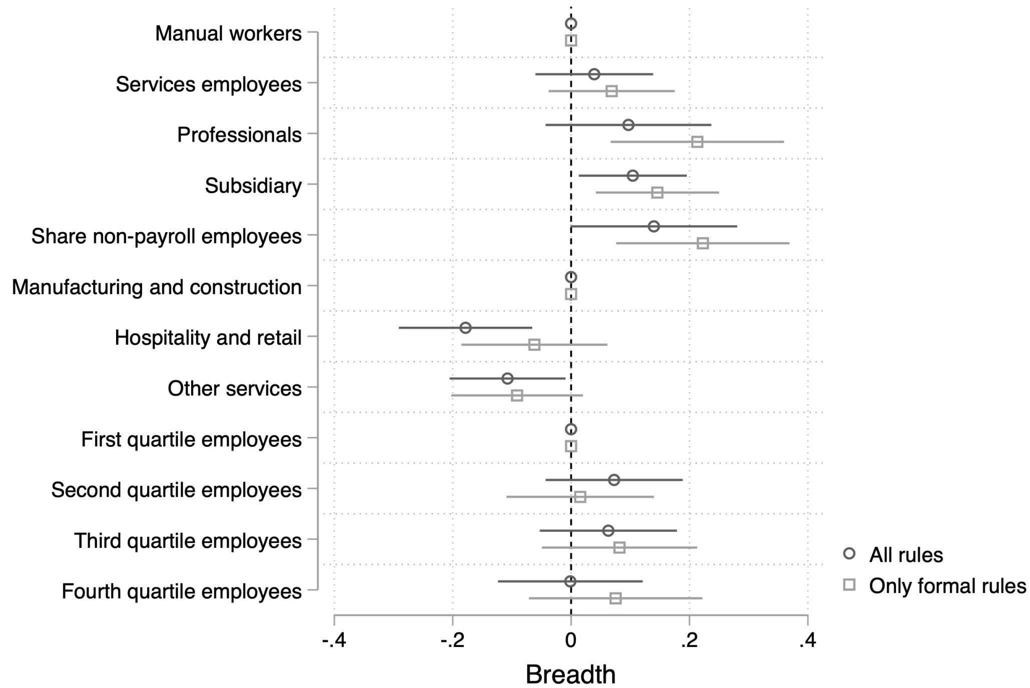
4.1 Rules and firm observables

In this section, I conduct a descriptive regression analysis to explore how firm characteristics relate to the breadth of the rulebook. The firm characteristics considered include firm size, sector, the type of employee position, and whether the firm is a subsidiary of a larger firm. Since each firm is observed in two scenarios, I obtain two measures of breadth per firm. To account for this, I run all regressions with standard errors clustered at the firm level. Specifically for firms $f \in \{1, 2, .., F\}$ and $s \in \{1, 2\}$ for outcomes y and covariates X

$$y_{fs} = \alpha + \boldsymbol{\beta} \boldsymbol{X}_f + \varepsilon_{fs}, \tag{2}$$

Figure 1 presents the results from these regressions. This figure shows three variables are correlated with the breadth of the rulebook used by firms: whether the main position being interviewed for is a professional position, whether the firm is a subsidiary of a larger firm, and the share of non pay-roll employees at the firm. This relationship appears to be driven by formal rather than

Figure 1: The relationship between firm observables and the breadth of the rulebook.



Notes This figure reports regression results from multivariate regressions of the breadth of the full and formal rulebook on various firm characteristics. These include employee type, categorised into manual workers, service employees, and professionals, with manual workers as the base category; the share of non-payroll employees; a subsidiary dummy variable where 1 denotes that the plant where the interview took place is a subsidiary of a larger entity; industry classification, which includes manufacturing and construction, retail and trade, and other service sectors such as tourism and education, with manufacturing as the base category; and firm size in terms of employment, included by quartile. Error bars represent 95% confidence intervals, with standard errors clustered at the firm level.

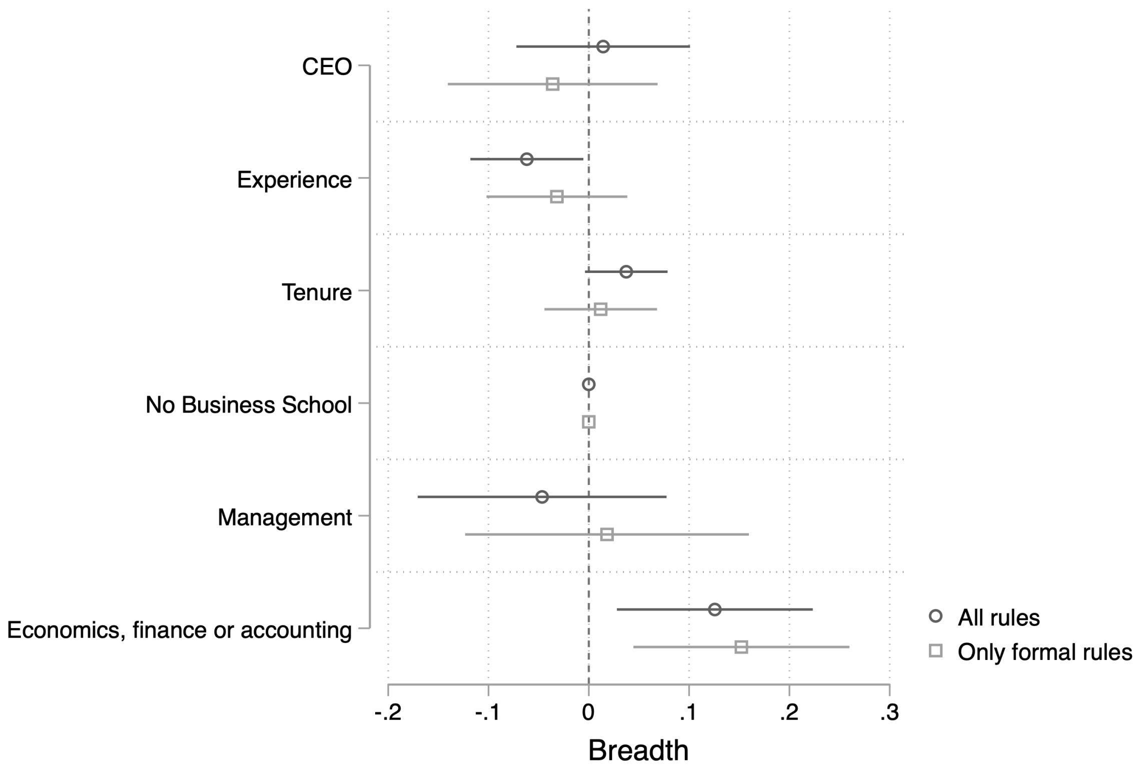
informal rules. These results suggest managers at different types of firms differ in the use of formal rules.[3]

4.2 Which managers use rules?

The following section focuses on whether different managers adopt different rules. Figure 2 summarises the results of regressions of the breadth of the rulebook on various manager characteristics. The results indicate that CEOs tend to use a narrower set of formal rules. Managers who studied economics, finance, or accounting at a business school tend to use broader rules, whereas managers who studied management at a business school do not differ significantly from those who did not attend a business school in their use of rules.[4]

[3] Appendix Table A.2 shows the results are comparable in a univariate framework.
[4] Appendix Table A.3 shows that the results are comparable in a univariate regression framework.

Figure 2: The relationship between manager characteristics and the breadth of the rulebook.



Notes This figure reports regression results from multivariate regressions of the breadth of the full and formal rulebook on various respondent characteristics. The CEO dummy indicates whether the respondent is the CEO or equivalent. Experience and tenure are respectively years of experience and years of experience at the firm, measured in units of five years. The management dummy indicates whether the respondent studied management at a business school. The economics dummy indicates if the manager studied economics, finance, or accounting at a business school. Error bars represent 95% confidence intervals.

In the survey, I elicit both a measure of the respondent's general trust and a measure of the respondent's trust in employees. Managers who are more trusting in general implement a more formal rule-based style of management, whereas managers who report higher trust in their employees implement a less rule-based management style. These effects are large and statistically significant. One interpretation is that higher trust in employees – potentially reflecting a well-functioning relational contract within the firm – reduces the perceived need for extensive rules. This finding aligns with Bloom et al. (2012), who show that higher trust within multinational organizations is associated with greater decentralization. Consistent with this, I observe a less rule-based environment when managers report greater trust in employees. Table 3 summarises these results.

Table 3: Rule breadth and trust

	(1) Breadth	(2) Breadth formal	(3) Breadth	(4) Breadth formal
General Trust	0.054	0.113***		
	(0.04)	(0.04)		
Employee Trust			-0.106***	-0.089**
			(0.04)	(0.04)
Constant	0.555***	0.208***	0.626***	0.295***
	(0.05)	(0.05)	(0.05)	(0.05)
Controls	Yes	Yes	Yes	Yes
N	536	536	560	560
N firms	269	269	281	281
R2	0.0518	0.0930	0.0561	0.0820

Notes This table displays the results from regressions of a self-reported measure of the manager's trust in general (General Trust) and their trust in employees (Employee trust). Both are a binary indicator for whether people in general/employees respectively can be trusted, with a higher value indicating greater trust. Standard errors are clustered at the firm-level, including two observations per firm. Standard errors are indicated in braces and significance is indicated as: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

I also ask respondents about their beliefs regarding good management through a series of hypothetical scenarios. For example, when hiring a candidate, should the decision be based on discretion or rules? Based on their responses, I create four indices: The *rules index* captures how respondents value rules over discretion. The *procedure index* measures the extent to which managers feel employees should follow procedures rather than show initiative. The *delegation index* captures whether managers believe decision-making should be delegated rather than involve the CEO. The *consistency index* assesses whether managers believe that current decisions set a precedent for future decisions. Appendix B details these questions and the construction of these indices. Each of these metrics is generated by combining a set of Likert-scale responses through inverse covariance weighting.

I regress the breadth of the formal rulebook on each of these indices. I find that managers who value rules, procedure, delegation, and consistency more tend to work in firms with broader formal rulebooks, showing that this measure is strongly correlated with managers' stated beliefs about good management. Table 4 summarises these results. This suggests that either these managers have sufficient influence to affect the formal rules of their firm, or that their beliefs regarding good management are, in part, shaped by operating in a rules-based environment.

Table 4: Rule breadth and managers' beliefs

	(1) Breadth formal	(2) Breadth formal	(3) Breadth formal	(4) Breadth formal
Rules	0.070^{***}			
	(0.02)			
Procedure		0.047^{**}		
		(0.02)		
Delegation			0.081^{***}	
			(0.02)	
Consistency				0.039^{*}
				(0.02)
Constant	0.287^{***}	0.264^{***}	0.268^{***}	0.275^{***}
	(0.05)	(0.05)	(0.05)	(0.05)
Controls	Yes	Yes	Yes	Yes
N	564	564	564	564
N firms	283	283	283	283
R2	0.0997	0.0850	0.1086	0.0806

Notes This table displays the results from regressions of a self-reported measure of the managers beliefs related to the importance of rules, procedure, delegation and consistency. These are all standardized indices from a set of questions on these topics, with higher values indicating a manager assigns greater value to rules, procedure, delegation and consistency respectively. Standard errors are clustered at the firm-level, including two observations per firm. Standard errors are indicated in braces and significance is indicated as: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.3 Rules and management practices

This section examines how rules-based management relates to other dimensions of firm management, focusing on structured management practices measured using MOPS-type questions (Bloom, Sadun, and Van Reenen, 2016) and HR practices. Results are reported in Table 5. To measure management quality, I construct scores as the weighted average of z-scores from MOPS-style questions, with weights determined by the inverse of the covariance matrix.

I find that the breadth of the full rulebook is only weakly related to management scores, whereas the breadth of the formal rulebook is more strongly positively correlated with management practices. One MOPS question addresses HR practices, asking whether promotions are based mainly on performance or on tenure and experience. Since higher values correspond to more performance-based promotions, firms with broader formal rulebooks use less performance-based promotion systems and make more rule-based promotion decisions. Notably, this HR item moves in the opposite direction to the rest of the management module: responses to the HR question are negatively related to responses to other management questions, and the management score excluding the HR item is strongly negatively correlated with the score based only on the HR item. One interpretation – perhaps highlighting a contextual difference from the high-performance work systems emphasized in Ichniowski, Shaw, and Prennushi (1997) – is that formalized promotion rules can raise productivity by improving consistency and perceived fairness in settings where high-frequency, high-quality measures of individual performance are harder to generate, verify, and act upon.

Table 5: Rule breadth and MOPS management practices

	(1)	(2)	(3)	(4)	(5)	(6)
		Breadth			Formal breadth	
MOPS Score	0.033**			0.050***		
	(0.02)			(0.02)		
MOPS Excluding HR		0.031*			0.068***	
		(0.02)			(0.02)	
MOPS Only HR			0.012			-0.038*
			(0.02)			(0.02)
Constant	0.563***	0.563***	0.561***	0.371***	0.371***	0.372***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
N	580	580	560	580	580	560
N firms	291	291	281	291	291	281
R2	0.0078	0.0068	0.0010	0.0156	0.0288	0.0089
Adj-R2	0.0061	0.0050	-0.0008	0.0139	0.0271	0.0071

Notes This table displays the results from regressions of the (formal) rule breadth on management practices as measured using MOPS. The MOPS score is the inverse covariance weighted average z-score of all responses to the management questions. MOPS excluding HR and MOPS HR only implements the same procedure including respectively all questions except the HR question, and only the HR question. Standard errors are clustered at the firm-level, including two observations per firm. Standard errors are indicated in braces and significance is indicated as: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A more detailed examination of the link between HR rules and HR practices reveals an interesting picture. Broadly speaking, firms with broader formal rulebooks appear to engage in more active HR management. Table 6 shows that HR managers in firms with broader formal rulebooks report a greater number of HR problems, and are more likely to cite challenges related to turnover, a lack of skills among potential employees, and low motivation. Although not shown here, these firms also exhibit greater labour market churn: relative to their workforce size, they have hired more workers (and expect to hire more) and have terminated a larger share of employees over the past 12 months. They spend more when hiring professional employees and also invest more when hiring other employees, and they grant pay rises more frequently. In recruitment, firms with broader rulebooks are more likely to advertise in a gazette or on Ethiojobs (and other vacancy websites), but are less likely to use recruitment agencies or advertise on university campuses. These relationships remain robust after controlling for other firm observables. Taken together, the evidence suggests that broader formal rulemaking is associated with more intensive HR activity.

Table 6: Rule breadth and HR management

	(1) Number of problems	(2) Motivation	(3) Turnover	(4) Lack of skills
Breadth formal	0.493***	0.131*	0.121*	0.242***
	(0.15)	(0.07)	(0.06)	(0.06)
Constant	1.358***	0.403***	0.468***	0.487***
	(0.16)	(0.07)	(0.07)	(0.07)
Controls	Yes	Yes	Yes	Yes
N	564	564	564	564
N firms	283	283	283	283
R2	0.0837	0.0378	0.0720	0.0896

Notes This table displays regression results linking the breadth of (formal) organisational rules to HR problems identified by managers. The first three dependent variables are dummy indicators for three HR issues: lack of motivation, high turnover, and skills shortage. Standard errors are clustered at the firm-level, including two observations per firm. Standard errors are indicated in braces and significance is indicated as: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5 Rules-based management and firm performance

To validate the measure of rules developed in this paper, I examine how rule-based management is correlated with firm performance, and whether it helps account for performance differences across otherwise similar enterprises. In this section, I study the association between the measures of rules developed in this paper and firms' sales, profits, and profit margins in three steps. I first document that rule breadth—especially *formal*, documented rules—is positively correlated with profitability, and most strongly with the profit margin. I then show that these correlations are robust to a rich set of firm observables, including a standard measure of management practices based on MOPS. Finally, I compare the incremental explanatory power of formal rule breadth and the MOPS score in accounting for cross-firm variation in performance, conditional on the same set of controls.

5.1 Rules and profitability

To examine whether rules-based management is associated with firm performance, I use profit and sales data collected from a subset of 189 firms in an exercise similar to that conducted in Bloom and Van Reenen (2007). Specifically, I estimate regressions of the form

$$y_{fs} = \alpha + \beta_1 \cdot \text{Breadth}_{fs} + \mathbf{X}'_f \boldsymbol{\beta}_2 + \varepsilon_{fs}, \tag{3}$$

where y_{fs} is firm f's performance measure in scenario s. The key regressor Breadth_{fs} is scaled from 0 to 1 and captures the share of managerial decisions made using rules in the hypothetical scenarios. The control vector $\mathbf{X}_f$ includes firm size, sector, the type of position discussed, whether the plant is a subsidiary, and (in some specifications) standard measures of management practices based on MOPS. Because I observe rule breadth twice per firm, I cluster standard errors at the firm level.

Table 7 reports results for three outcomes: monthly sales, monthly profit, and the profit margin (profit over sales). Across specifications, rule breadth is not systematically related to sales conditional on observables. In contrast, rule breadth is positively associated with profitability. Interpreting the coefficient on a 0–1 index, moving from no rule use to full rule use is associated with substantially higher monthly profit and a higher profit margin.

The results are particularly pronounced for formal rules. Columns (4)–(6) of Table 7 show that formal rule breadth is not significantly related to sales, but is positively associated with both profit and profit margins. A one-standard-deviation increase in formal rule breadth is associated with approximately 4 million ETB higher monthly profit – roughly one-third of the sample mean – and a 2.5 percentage point higher profit-to-sales ratio ($p < 0.01$).

Table 7: Organisational Rules and Firm Performance

	(1) Revenue	(2) Profit	(3) Profit/Sales	(4) Revenue	(5) Profit	(6) Profit/Sales
Breadth	-7.042	11.197*	0.065***			
	(60.60)	(5.73)	(0.02)			
Breadth formal				58.890	10.893**	0.063***
				(64.83)	(5.01)	(0.02)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. var.	236.527	12.735	0.091	236.527	12.735	0.091
N	377	365	331	377	365	331
N firms	189	183	166	189	183	166
R2	0.466	0.280	0.077	0.467	0.280	0.076
Adj-R2	0.454	0.264	0.054	0.456	0.264	0.053

Notes This table displays the relationship between firm performance and rule breadth. The breadth variable captures the span of formal rules and informal rules taken together, whereas breadth formal only includes the formal rules. The three dependent variables included are sales in the past month in millions of Birr, profit in the past month in millions of Birr, and the ratio of profit over sales. These variables are all winsorised at the 95th percentile. The included controls are the number of employees, sector, the type of position, management practices, and whether the firm is a subsidiary. All standard errors are clustered at the firm level. Figure 1 details these variables; the number of employees is incorporated as a continuous variable rather than by quartile. Standard errors are indicated in braces and significance is indicated as: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.2 Robustness to controls

A natural concern is that firms with more formal rules differ systematically in other dimensions—such as size, industry, or management practices—and that these differences account for the correlation with profitability. To assess sensitivity to selection on observables, focusing on formal rules, Table 8 progressively adds blocks of controls.

Panels A and B suggest that the positive association between formal rule breadth and profitability is stable across specifications: the point estimates remain positive and statistically significant as additional covariates are included. Panel C shows a similarly positive relationship for the profit margin. Taken together, these results indicate that the correlation between formal rule use and profitability is not driven solely by observable differences in firm size, industry, or measured management practices.

Table 8: Organisational Rules and Firm Performance

	Panel A: Revenue					
	(1) Revenue	(2) Revenue	(3) Revenue	(4) Revenue	(5) Revenue	(6) Revenue
Breadth formal	162.495*	86.110	95.755	63.263	58.492	58.424
	(86.25)	(64.44)	(65.17)	(60.98)	(61.91)	(64.70)
Unc. mean dep. var.	236.527	236.527	236.527	236.527	236.527	236.527
N	387	387	387	387	377	377
N firms	194	194	194	194	189	189
Adj-R2	0.014	0.444	0.447	0.457	0.455	0.456

	Panel B: Profit					
	(1) Profit	(2) Profit	(3) Profit	(4) Profit	(5) Profit	(6) Profit
Breadth formal	16.240***	11.925**	12.432***	13.491**	11.601**	11.079**
	(4.87)	(4.66)	(4.71)	(5.24)	(5.09)	(5.01)
Unc. mean dep. var.	12.735	12.735	12.735	12.735	12.735	12.735
N	375	375	375	375	365	365
N firms	188	188	188	188	183	183
Adj-R2	0.033	0.227	0.229	0.230	0.256	0.264

	Panel C: Profit over sales					
	(1) Profit/Sales	(2) Profit/Sales	(3) Profit/Sales	(4) Profit/Sales	(5) Profit/Sales	(6) Profit/Sales
Breadth formal	0.055**	0.057**	0.061**	0.062***	0.059***	0.064***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Unc. mean dep. var.	0.091	0.091	0.091	0.091	0.091	0.091
N	339	339	339	339	331	331
N firms	170	170	170	170	166	166
Adj-R2	0.019	0.018	0.020	0.017	0.040	0.051
Firm Size	No	Yes	Yes	Yes	Yes	Yes
Management Practices	No	No	Yes	Yes	Yes	Yes
Subsidiary	No	No	No	Yes	Yes	Yes
Position	No	No	No	No	Yes	Yes
Industry	No	No	No	No	No	Yes

Notes This table examines the robustness of the relationship between firm revenue, profitability, and profit over sales with formal rule breadth. All three dependent variables are winsorised at the 95th percentile. The included controls are the number of employees, sector, the type of position, management practices, and whether the firm is a subsidiary. Standard errors, clustered at the firm-level, are indicated in braces and significance is indicated as: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.3 Explaining performance differences between firms

Finally, I assess whether this paper's measure of rules-based management contains predictive information about firm performance beyond standard measures of management quality. Table 9 compares specifications that include (i) controls only, (ii) controls plus the MOPS score, (iii) controls plus formal rule breadth, and (iv) controls plus both measures.

For sales, neither formal rule breadth nor the MOPS score materially improves model fit relative to the baseline specification with controls. For profits and profit margins, however, adding formal rule breadth increases model fit, whereas adding the MOPS score provides little additional explanatory power and does not improve fit conditional on formal rule breadth.

Table 9: Organisational Rules and Firm Performance

	Panel A: Sales			
	(1) Revenue	(2) Revenue	(3) Revenue	(4) Revenue
MOPS Score		-27.585		-30.304
		(21.82)		(22.35)
Breadth formal			49.581	58.890
			(64.70)	(64.83)
	(43.71)	(45.55)	(44.15)	(46.62)
Mean dep. var.	236.527	236.527	236.527	236.527
N	377	377	377	377
N firms	189	189	189	189
R2	0.4626	0.4655	0.4640	0.4674
Adj-R2	0.4539	0.4554	0.4539	0.4559
	Panel B: Profit			
	(1) Profit	(2) Profit	(3) Profit	(4) Profit
MOPS Score (excl. HR)		-0.138		-0.681
		(1.58)		(1.60)
Breadth formal			10.414**	10.626**
			(5.01)	(4.99)
	(3.34)	(3.39)	(3.50)	(3.56)
Mean dep. var.	12.735	12.735	12.735	12.735
N	365	365	365	365
N firms	183	183	183	183
R2	0.2638	0.2638	0.2771	0.2774
Adj-R2	0.2515	0.2494	0.2629	0.2612
	Panel C: Profit over sales			
	(1) Profit/Sales	(2) Profit/Sales	(3) Profit/Sales	(4) Profit/Sales
MOPS Score		-0.004		-0.009
		(0.01)		(0.01)
Breadth formal			0.059***	0.063***
			(0.02)	(0.02)
Mean dep. var.	0.091	0.091	0.091	0.091
N	331	331	331	331
N firms	166	166	166	166
R2	0.0506	0.0514	0.0737	0.0764
Adj-R2	0.0331	0.0308	0.0536	0.0534
Controls	Yes	Yes	Yes	Yes

Notes This table compares the predictive content of formal rule breadth and the MOPS score for sales, profit and profit over sales. All standard errors are clustered at the firm level. The included controls are the number of employees, sector, the type of position, management practices, and whether the firm is a subsidiary. Standard errors, clustered at the firm-level, are indicated in braces and significance is indicated as: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6 Rules and Resilience

Existing research establishes a connection between a firm's resilience to economic shocks and the autonomy granted to its managers and employees (Aghion et al., 2021; Li et al., 2022). In particular, Li et al. (2022) develops a theoretical model in which economic shocks cause relational contracts to become more rule-based, and these new rules then reduce both long-term performance and resilience to future shocks. As the survey data was collected at the end of 2021, I included questions on the effect of COVID-19 on firms to test this hypothesis.

The results present a puzzle: while Section 5 showed that rules-based management is associated with higher profitability, I find that such firms were also *more vulnerable* to the COVID-19 shock. This pattern is consistent with Li et al.'s (2022) theoretical prediction that rules reduce organizational flexibility, but it raises an important identification question: does adopting rules *cause* vulnerability, or do firms adopt rules *in response to* negative shocks?

6.1 Rules and the impact of COVID-19

I regress four measures of COVID-19 impact on the breadth of the organizational rulebook: (1) whether cash flow decreased due to COVID-19, (2) whether the firm had to postpone or cancel purchases due to the pandemic, (3) whether the firm experienced financial problems, and (4) whether profit decreased in the financial year following the pandemic compared to the financial year ending shortly after the first lockdown. Throughout, I include the same set of controls as when assessing firm profitability: number of employees, sector, type of position, management practices, and whether the firm is a subsidiary.

Panel A of Table 10 reports results for the full set of organizational rules (both formal and informal). Firms with broader rulebooks were significantly more likely to experience negative COVID-19 effects across all four outcomes. Specifically, a one-unit increase in rule breadth is associated with a 15.0 percentage point increase in the probability of decreased cash flow ($p < 0.01$), an 18.3 percentage point increase in cancelled purchases ($p < 0.01$), a 12.1 percentage point increase in financial problems ($p < 0.10$), and a 2.4 percentage point increase in falling profits. These are economically substantial effects.

6.2 Distinguishing Causation from Response

The correlation between rules and COVID-19 vulnerability admits two interpretations. First, rules may *cause* reduced resilience by limiting organizational flexibility during shocks, as Li et al. (2022) theorizes. Second, firms may *adopt rules in response to* experiencing negative shocks, perhaps as managers attempt to regain control during turbulent periods. I implement three tests to distinguish between these possibilities.

The first test exploits the distinction between formal and informal rules. Formal rules—written policies, explicit procedures—are relatively costly to implement and change. Informal rules—norms, conventions, implicit expectations—can be adopted more quickly and with less deliberation. If the correlation between rules and COVID-19 impacts reflects a *causal* effect of rules on vulnerability, we would expect both formal and informal rules to show similar patterns. If instead the correlation reflects firms *adopting rules in response to* shocks, we would expect the effect to be concentrated in informal rules, which can be adopted more rapidly.

Panel B of Table 10 reports results using only formal rules. The coefficients are substantially smaller and statistically insignificant across all four outcomes: 0.034 for decreased cash flow (compared to 0.150 in Panel A), 0.130 for cancelled purchases (compared to 0.183), 0.024 for financial problems (compared to 0.121), and -0.064 for falling profits (compared to 0.024). This asymmetry suggests that the relationship observed in Panel A is driven primarily by *informal* rules. Since informal rules can be adopted quickly, this pattern is more consistent with rules being a *response to* shocks rather than a *cause* of vulnerability.

Table 10: Relationship between breadth of the rulebook and firm resilience.

Panel A: All rules and resilience

	(1) Decreased cash-flow	(2) Cancelled purchases	(3) Financial problems	(4) Fall profit
Breadth	0.156**	0.183***	0.121*	0.024
	(0.06)	(0.05)	(0.07)	(0.09)
Constant	0.766***	0.137**	0.765***	0.522***
	(0.07)	(0.06)	(0.07)	(0.10)
Controls	Yes	Yes	Yes	Yes
N	564	564	564	361
N firms	283	283	283	181
R2	0.0773	0.0502	0.0757	0.0467

Panel B: Formal rules and resilience

	(1) Decreased cash-flow	(2) Cancelled purchases	(3) Financial problems	(4) Fall profit
Breadth formal	0.034	0.130**	0.034	-0.064
	(0.06)	(0.05)	(0.06)	(0.09)
Constant	0.848***	0.210***	0.827***	0.549***
	(0.05)	(0.06)	(0.06)	(0.09)
Controls	Yes	Yes	Yes	Yes
N	564	564	564	361
N firms	283	283	283	181
R2	0.0608	0.0385	0.0677	0.0488

Notes This table examines the relationship between the breadth of the identified rules and the effects of Covid-19 restrictions on the firm. Decrease cash flow, postpone purchases and financial problems equal one if the manager indicates they had any of these issues due to Covid-19 restrictions. The indicator fall profit equals one if the firm reports lower profit in the 2020-2021 financial year than the 2019-2020 financial year. The included controls are the number of employees, sector, the type of position, management practices, and whether the firm is a subsidiary. Standard errors, clustered at the firm level, are indicated in braces and significance is indicated as: * $p < 0.10$, **, $p < 0.05$, *** $p < 0.01$.

The second test exploits panel data on a subset of firms surveyed in both 2017 and 2021. While I do not have data on rules for 2017, I use the adoption of rules-based management practices (measured using the MOPS survey instrument) as a proxy. If rules-based management *causes* vulnerability, we would expect *pre-COVID* (2017) management practices to predict COVID-19 impacts. If instead firms adopt rules-based approaches *in response to* shocks, only *contemporaneous* (2021) practices should correlate with outcomes.

Table 11 reports results from this exercise. Panel A shows that contemporaneous (2021) rules-based management practices are positively correlated with financial problems (0.074, $p < 0.01$) and falling profits (0.065, $p < 0.10$). However, Panel B shows that *past* (2017) management practices have no significant relationship with any COVID-19 outcome. The coefficients are close to zero and precisely estimated: -0.023 for decreased cash flow, 0.026 for cancelled purchases, -0.023 for financial problems, and -0.039 for falling profits. The absence of any predictive power from pre-COVID management practices, combined with significant effects from contemporaneous practices, supports the interpretation that rules-based management is adopted in response to shocks rather than causing vulnerability.

Taken together, these results suggest that the positive correlation between rules and COVID-19 vulnerability reflects firms adopting rules-based management *in response to* negative shocks, consistent with Li et al.'s (2022) theoretical framework. The evidence against a causal interpretation comes from two sources: (1) the concentration of effects in informal rather than formal rules, and (2) the absence of predictive power from pre-shock management practices. These findings have implications for understanding organizational adaptation during crises: rather than rules causing rigidity, the adoption of rules may be a symptom of firms struggling to cope with disruption.

Table 11: Relationship between management practices and firm resilience.

Panel A: Current management practices and resilience

	(1) Decreased cash-flow	(2) Cancelled purchases	(3) Financial problems	(4) Fall profit
MOPS HR Score	-0.024	0.005	0.074***	0.065*
	(0.02)	(0.03)	(0.03)	(0.04)
Constant	0.751***	0.210***	0.644***	0.476***
	(0.03)	(0.02)	(0.03)	(0.04)
N	281	281	281	179
N firms	281	281	281	179
R2	0.0032	0.0002	0.0237	0.0188

Panel B: Past management practices and resilience

	(1) Decreased cash-flow	(2) Cancelled purchases	(3) Financial problems	(4) Fall profit
MOPS HR Score	-0.023	0.026	-0.023	-0.039
	(0.03)	(0.03)	(0.04)	(0.05)
Constant	0.746***	0.197***	0.609***	0.438***
	(0.03)	(0.03)	(0.04)	(0.05)
N	173	173	174	110
N firms	173	173	174	110
R2	0.0027	0.0042	0.0021	0.0068

Notes This table examines the relationship between HR management practices and the effects of Covid-19 restrictions on the firm. Decrease cash flow, postpone purchases and financial problems equal one if the manager indicates they had any of these issues due to Covid-19 restrictions. The indicator fall profit equals one if the firm reports lower profit in the 2020-2021 financial year than the 2019-2020 financial year. Standard errors, clustered at the firm level, are indicated in braces and significance is indicated as: * $p < 0.10$, **, $p < 0.05$, *** $p < 0.01$.

7 Assessing the performance of the sampling algorithm

This section assesses the performance of the Bayesian adaptive algorithm through three analyses: (1) analysing data collected from firm managers in the pilot, (2) using simulated data to identify settings where the algorithm performs well, and (3) comparing the methodology to that of Chapman et al. (2024) in measuring the parameters of a utility function.

7.1 Measuring managerial decisions using adaptive sampling

I assess whether the adaptive questionnaire outperforms random question selection by using data from a different sample of 198 managers, drawn from the same population. Each manager made decisions for eight randomly selected cases from the scenario landscape. I use this data to test the adaptive algorithm against random sampling. Appendix D reports the estimated parameters for the GP based on this sample used in the analysis and data collection.

To compare the performance of the adaptive algorithm and random sampling, I use training data to conduct an exercise where each respondent's eight observations are split into seven training cases and one test case repeatedly. I then compare the performance of adaptive versus random sampling in predicting the value of that test case after each sampling step. The adaptive algorithm's accuracy is 6% higher for the pay rise scenario and 5% higher for the hiring scenario after three questions. Figure 7.1 depicts the results from an OLS regression comparing random and adaptive sampling.

Table 12 depicts the performance of adaptive relative to random sampling. The top two rows of results depict the efficiency ratio of the adaptive sampling algorithm. The sampling algorithm increases the learning rate by 60% to 80%.

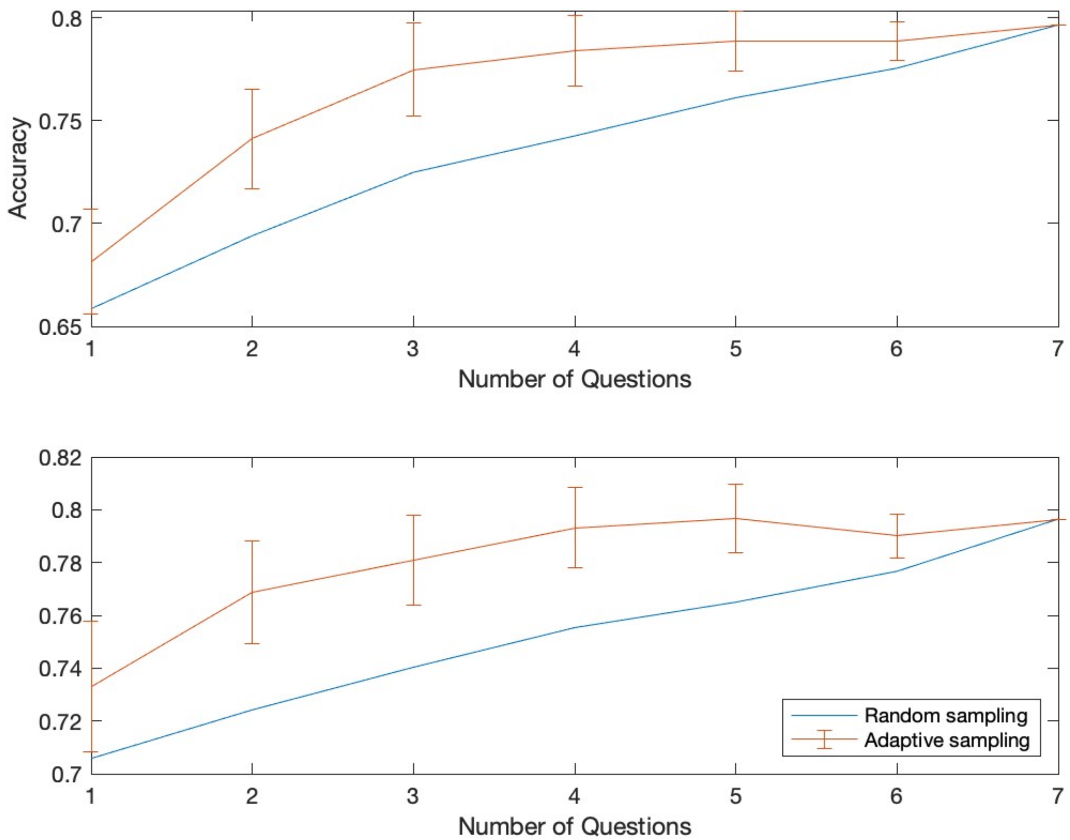
However, in this exercise the adaptive sampling algorithm is heavily constrained, as it can only choose from the seven randomly sampled questions rather than the full set of 128 cases. To evaluate the algorithm's potential without this constraint, I follow the literature on optimal experimental design (Chapman et al., 2024). Specifically, I train the model using the full dataset, to then simulate individuals for whom we have observed the full set of cases using the trained prior and the observed training data to generate a set of decisions for all cases. The second two rows of results in Table 12 show the adaptive algorithm now increases the rate of learning by 50 to 300% compared to randomly sampling questions. As described in section 7.2, the algorithm does particularly well when there is a lot of structure to exploit, which is the case in the pay rise scenario.

This section shows the algorithm significantly increases our rate of learning relative to randomly sampling questions. Based on these results, I would expect that the accuracy after eight adaptively sampled questions is approximately the same as after 15 to 20 randomly sampled questions.

7.2 Analysis using simulated data

In this section, I study the performance of the Bayesian adaptive questionnaire using simulated data. These simulations help identify situations where the adaptive algorithm performs particularly well. I simulate data from various Gaussian Processes (GPs) and compare the performance of the adaptive sampling algorithm to random sampling. The simulations show that the adaptive algorithm significantly increases the rate of learning, especially when it can exploit the structure in the relationship between outcomes Y and characteristics X. This is evident when different characteristics contribute differently to the covariance and when the mean function provides information

Figure 3: Assessing the performance of the adaptive algorithm using training data



Notes This figure compares the performance of randomly sampling observations (the blue line) with sampling them adaptively (in red). The set of responses is repeatedly split in a training set (7 cases) and test set (1 case) for each respondent to implement this. The confidence bands for the performance of the adaptive sampling algorithm are based on standard errors clustered at the (simulated) individual level. The accuracy is the average probability that the algorithm makes a correct prediction for the test case.

Table 12: Efficiency Ratio training sample

	Constrained Efficiency Ratio		
Scenario	2 Questions	3 Questions	4 Questions
Pay rise scenario	1.57	1.66	1.73
Hiring scenario	1.78	1.76	1.77
N	198	198	198
	Unconstrained Efficiency Ratio		
Scenario	5 Questions	10 Questions	15 Questions
Pay rise scenario	1.57	2.57	4.12
Hiring scenario	1.56	2.17	2.59

Notes Table 12 depicts the efficiency ratio of the adaptive sampling algorithm relative to random sampling using the training data. For each respondent, the answers are repeatedly randomly split in a training set (7 cases) and test set (1 case). The adaptive and random sampling algorithm are applied to the training set. I generate a benchmark based on the average accuracy of the two algorithms after 2, 3 and 4 questions. I then determine the first point where each of these algorithms achieves this accuracy to calculate, respectively N and N^* for random and adaptive sampling. I calculate the efficiency gains as the mean of N over the mean of N^* to create a symmetric measure. The constrained efficiency ratio uses only the real data. The unconstrained efficiency ratio uses the observed data to simulate the full set of observations. To do so, I train the hyperparameters of the Gaussian Process, specifically a linear mean function and automatic relevance detection square exponential kernel. I use this to generate data based on the observed data for each individual and this prior trained on the full training sample. In predicting managers' decisions using real data, the adaptive algorithms achieves 60-80% faster learning than random sampling. In the simulations using the data that was simulated using the real data the adaptive sampling algorithm massively outperforms random sampling.

Table 13: The eight data-generating processes used for simulations

	No Mean Function	Mean Function
ℓ_1 equals 1	DGP 1	DGP 5
ℓ_1 is very low	DGP 2	DGP 6
ℓ_1 is high	DGP 3	DGP 7
ℓ_1 and ℓ_2 are high	DGP 4	DGP 8

Notes: The data-generating processes all have six binary characteristics that determine a binary outcome. The covariance is modeled using an automatic relevance detection squared exponential kernel. Unless otherwise noted, the length-scale parameters for the remaining characteristics are set to 1. The mean function is a simple linear mean with some random noise added.

about predictability. Additionally, the more questions we ask a respondent, the better the algorithm becomes at selecting informative questions based on prior responses.

To simulate the data, I use a squared exponential kernel with automatic relevance detection (ARD). In this specification, each case $\mathbf{x}_i$ has D binary characteristics, and ℓ_d is the length-scale parameter for characteristic d. In this kernel, the covariance between two cases decreases exponentially with the squared distance between their characteristics, where the length-scale parameter ℓ_d controls sensitivity to differences in each characteristic.

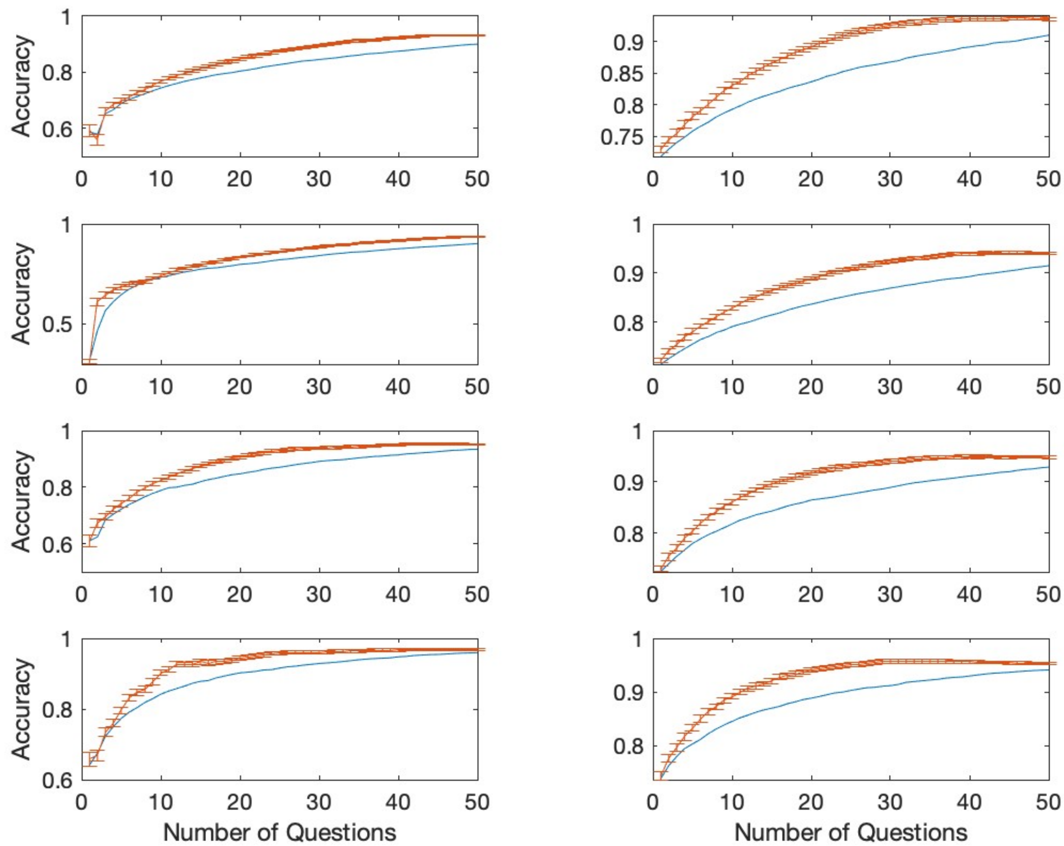
$$k(\mathbf{x}_p, \mathbf{x}_q) = \exp\left(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_{pd}-x_{qd})^2}{\ell_d^2}\right) \tag{4}$$

Each data-generating process (DGP) has six binary characteristics. Generally, ℓ_d is set to 1, except where explicitly stated otherwise. For process 1, ℓ_d is set to 1 for all characteristics. In process 2, I set ℓ_1 to a very small value, which means that the covariance between cases that differ in characteristic 1 drops off to nearly zero, effectively treating those cases as uncorrelated in this dimension. For processes 3 and 4, I set one and two length-scale parameters respectively to high values, meaning that differences in those specific characteristics have little effect on the covariance between cases. Processes 5-8 use the same four combinations of length-scale parameters, but with a linear mean function added to the Gaussian Process. Table 13 summarizes these processes.

In Figure 4, I compare the adaptive Bayesian questionnaire to randomly sampling a sequence of questions. This figure shows that as the structure on the problem increases, the adaptive sampling algorithm can leverage this information to further outperform random sampling. All these problems have sufficient structure for adaptive sampling to significantly outperform random sampling. Detailed regression results relating to the performance gains can be found in Table A.1 in Appendix A.

While the absolute gains in accuracy may appear modest, a more meaningful metric is the efficiency ratio, calculated as the ratio of the number of questions required by random sampling (N) to achieve the same level of information as the optimized adaptive sampling (N^*), following

Figure 4: Assessing the performance of the adaptive algorithm using simulated data



Notes: This figure compares the performance of randomly sampling observations (the blue line) with sampling them adaptively (in red). The confidence bands for the performance of the adaptive sampling algorithm are based on standard errors clustered at the (simulated) individual level. The accuracy is the average probability that the algorithm makes a correct prediction for each of the decision problems. The specific DGPs used are described in Table 13.

(Chapman et al., 2024). Table 14 presents these efficiency ratios:

Table 14: Efficiency ratio of adaptive versus random sampling

DGP	Efficiency Ratio after 5 Questions	10 Questions	15 Questions
DGP 1	1.16	1.26	1.33
DGP 2	1.37	1.23	1.28
DGP 3	1.48	1.38	1.48
DGP 4	1.30	1.59	1.59
DGP 5	1.78	1.80	2.01
DGP 6	1.76	1.84	1.89
DGP 7	1.87	2.05	2.10
DGP 8	1.80	2.13	2.19

Notes This table reports the efficiency ratio of the adaptive sampling algorithm relative to random sampling. For each randomly generated individual, the adaptive and random sampling algorithm are applied. I generate a benchmark based on the average accuracy of the two algorithms after 5, 10 and 15 questions. I then determine the first point where each of these algorithms achieves this accuracy to calculate N and N^*. I calculate the efficiency gains as the mean of N over the mean of N^* to create a symmetric measure. Here, the adaptive algorithms achieves 15-120% faster learning than random sampling depending on the structure of the problem.

The results indicate that adaptive sampling requires significantly fewer questions to achieve the same level of accuracy as random sampling. The efficiency gains range from 15% to 120%, depending on the DGP. The gains are larger when the problem has more structure that the adaptive algorithm can exploit. One might argue that random sampling is a naive benchmark. However, alternative strategies, such as sampling cases that are very different from each other, may not perform better. In fact, such strategies can worsen performance compared to random sampling, as shown in Appendix Figure A.1.

These simulations demonstrate that the Bayesian adaptive questionnaire significantly outperforms random sampling in identifying informative questions, especially when the underlying decision process has exploitable structure. This efficiency is valuable in practical applications where the number of questions must be limited to reduce respondent burden.

7.3 Measuring Preferences

To further assess the performance and flexibility of the Bayesian adaptive questionnaire, I compare it to the Dynamic Optimisation of Sequential Experiments (DOSE) algorithm developed by Chapman et al. (2024). DOSE selects questions by maximising expected information gain about preference parameters (measured via Kullback–Leibler divergence). In their application, respondents choose between a lottery and a certain payoff.

Let the lottery pay $x \geq 0$ in the gain state and $z < 0$ in the loss state, and let the alternative be a sure payoff y. Utility over outcomes w is given by

$$u(w) = \begin{cases} w^{\rho}, & \text{if } w \geq 0, \\ -\lambda(-w)^{\rho}, & \text{if } w < 0, \end{cases} \tag{5}$$

where ρ captures curvature (risk aversion) and λ captures loss aversion. Choices are then generated from a stochastic choice rule with parameter μ governing noise.

To implement the Bayesian adaptive questionnaire in this context, I incorporate the same model structure. After each question, I update the posterior distribution of the parameters based on the respondent's answer and use this updated distribution to inform the selection of the next question. The mean function and kernel are updated accordingly.

The set of possible questions involves lotteries where the certain payoff y is set to zero, and the lottery payoffs x (positive) range from 0.25 to 5 in increments of 0.25, while z (negative) ranges from -5 to 0.5 in increments of 0.5, resulting in 200 possible questions. This setup mirrors that of Chapman et al. (2024).

I simulate data for 200 individuals and compare the performance of the Bayesian adaptive questionnaire to the DOSE algorithm in estimating the three parameters. The results show no significant difference between the two methods. After 30 questions, the Bayesian adaptive questionnaire marginally outperforms DOSE in estimating two of the three parameters. After 100 questions, DOSE performs slightly better in estimating two parameters. However, these differences are not statistically significant.

This comparison shows that the Bayesian adaptive questionnaire matches the performance of specialized algorithms like DOSE in this setting. The key advantage of the Bayesian adaptive questionnaire is its flexibility; it can be applied to a wide range of problems without substantial modifications. This makes it a valuable tool for various applications in experimental design and survey methodology, beyond the specific context of measuring preferences. This comes at the cost of increased computational requirements.

8 Discussion

Explaining persistent performance differences between seemingly similar enterprises is fundamentally important for designing and targeting policy. The question of what managers do, and what makes managers effective, is of first-order importance to firms in low-income countries. In the development of modern production, rules were fundamental in unlocking economies of scale in U.S. manufacturing (Taylor, 1911); yet, we know very little about how managers use rules today. We know that firms in sub-Saharan African countries reap limited benefits from economies of scale (Bassi et al., 2023), and that firms in low-income countries face high costs of delegation (Akcigit et al., 2021a; Bloom et al., 2013). The methodology developed here allows us to measure a dimension of organisational rules that appears to matter for firm performance, enabling the study of a wide range of questions about managerial practices and their impact on firm outcomes.

I argue that we know so little about rules because measuring such a high-dimensional construct is difficult using traditional economic methodologies. To address this challenge, I have developed a novel Bayesian adaptive questionnaire to measure organisational rules. This tool efficiently manages the high dimensionality by sampling the most informative questions. I implemented this methodology to survey a set of large Ethiopian firms. The results show that having a broader formal rulebook is positively correlated with firm performance and can help explain heterogeneity in performance between firms. These findings are robust to the inclusion of additional firm observables as explanatory variables. However, aligned with theoretical results from Li et al. (2022), this increased performance appears to come at the cost of greater vulnerability to economic shocks. Firms that have more rules perform better but report being more negatively affected by the COVID-19 restrictions implemented in Ethiopia in 2020.

This suggests that implementing formal rules might enhance the performance of relational contracts within firms but comes at the cost of increased vulnerability to changes in the economic environment. This aligns with the findings of Aghion et al. (2021), which show that reduced autonomy—and thus flexibility—at the plant level diminishes firms' resilience to economic turbulence. Crucially, the results in this paper demonstrate that organisational rules are a significant factor in explaining heterogeneity in performance across firms.

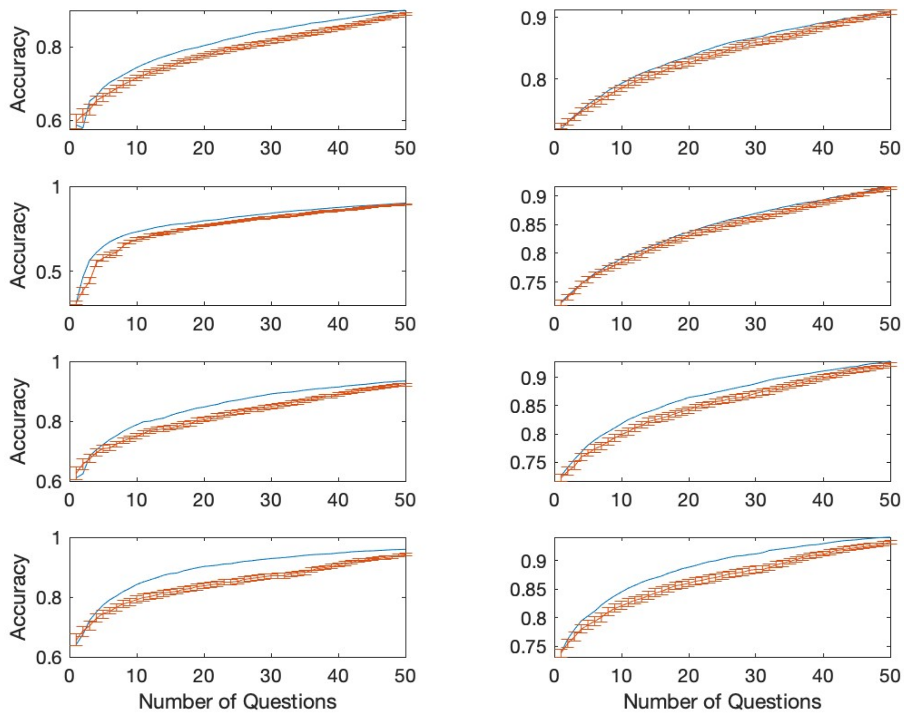
The positive association between formal rules and profitability is consistent with several theoretical channels: rules reduce delegation costs by allowing managers to extend their span of control without sacrificing consistency (Aghion and Tirole, 1997; Bloom et al., 2012); they support coordination among employees who must act interdependently without continuous communication (Gibbons and Henderson, 2012; March and Simon, 1958); and they may enhance perceived procedural fairness. These mechanisms may be particularly salient in Ethiopia, where delegation costs are known to constrain firm growth (Akcigit et al., 2021a; Bassi et al., 2023). Complementary evidence from Abebe et al. (2025) reinforces this interpretation: Ethiopian firms consistently prefer rule-based managerial styles when hiring, suggesting that such approaches are valued precisely because they facilitate delegation and coordination. Our cross-sectional design cannot isolate the relative importance of these channels, but the robustness of the profitability relationship suggests that organisational rules capture something meaningful about how firms operate.

Understanding how managers use rules is fundamentally important, particularly in low-income countries where managerial span of control appears to be a key constraint (Bassi et al., 2023; Bloom, Mahajan, McKenzie, and Roberts, 2010). The development of a feasible tool to measure rules opens up a number of avenues for policy-relevant research. First, using experimental exogenous variation to promote the implementation of formal rules could assess the causal link between rules and firm performance, with a strong focus on heterogeneity in treatment effects. Second, studying the shared understanding and enforcement of rules within a firm would provide insights into internal dynamics. Third, intersecting with the first two, examining how rules affect a firm's ability to react to changes—and whether the lower resilience to COVID-19 observed in this paper reflects a reduction in flexibility—would be valuable. Throughout this research, focusing on the theoretical modelling of rules in relational contracts is important to understand how rules affect outcomes.

The second contribution of this paper is the development of a novel method of questionnaire design. This method is broadly applicable and outperforms alternative generic sampling algorithms. Compared to random sampling, this algorithm doubles the rate of learning. The application presented in this paper demonstrates that the method can be feasibly implemented in a field setting. This approach has broad future applications in survey design where asking questions is costly. For example, applying this method to phone surveys is a potential next step. This Bayesian adaptive questionnaire methodology has the potential to significantly reduce the cost, both for interviewers and respondents, of collecting data for researchers and policymakers.

A Performance of the adaptive sampling algorithm

Figure A.1: Assessing the performance of adaptive versus dispersed sampling



Notes: This figure compares the performance of sampling a dispersed set of observations (the blue line) with sampling them adaptively (in red). The confidence bands for the performance of the adaptive sampling algorithm are based on standard errors clustered at the (simulated) individual level. The accuracy is the average probability that the algorithm makes a correct prediction for each of the decision problems. The dispersed questions are sampled by determining at each step which unobserved case has the fewest shared characteristics with the set of observed cases. The specific DGPs used are described in Table 13.

Table A.1: Gains in accuracy from adaptive sampling

DGP		(1) Step 5	(2) Step 10	(3) Step 15	(4) Step 25
1	Average increase due to adaptive sampling	0.00431	0.00928*	0.0345***	0.0478***
		(0.0107)	(0.00497)	(0.00271)	(0.00235)
	Average accuracy for random sampling	0.588***	0.689***	0.780***	0.827***
		(0.00885)	(0.00559)	(0.00385)	(0.00286)
2	Average increase due to adaptive sampling	-0.00656	0.0388***	0.0211***	0.0406***
		(0.00584)	(0.00601)	(0.00299)	(0.00229)
	Average accuracy for random sampling	0.315***	0.645***	0.773***	0.820***
		(0.00453)	(0.00619)	(0.00356)	(0.00275)
3	Average increase due to adaptive sampling	-0.00119	0.0186***	0.0550***	0.0589***
		(0.0111)	(0.00509)	(0.00326)	(0.00269)
	Average accuracy for random sampling	0.610***	0.722***	0.819***	0.870***
		(0.00966)	(0.00599)	(0.00389)	(0.00296)
4	Average increase due to adaptive sampling	0.0167	0.0225***	0.0554***	0.0423***
		(0.0104)	(0.00508)	(0.00347)	(0.00258)
	Average accuracy for random sampling	0.642***	0.773***	0.878***	0.916***
		(0.0103)	(0.00618)	(0.00373)	(0.00279)
5	Average increase due to adaptive sampling	0.0109***	0.0238***	0.0485***	0.0575***
		(0.00250)	(0.00289)	(0.00227)	(0.00204)
	Average accuracy for random sampling	0.719***	0.759***	0.818***	0.856***
		(0.00553)	(0.00482)	(0.00350)	(0.00271)
6	Average increase due to adaptive sampling	0.00797***	0.0265***	0.0490***	0.0533***
		(0.00200)	(0.00270)	(0.00215)	(0.00185)
	Average accuracy for random sampling	0.715***	0.756***	0.815***	0.855***
		(0.00522)	(0.00477)	(0.00375)	(0.00291)
7	Average increase due to adaptive sampling	0.00719	0.0352***	0.0502***	0.0514***
		(0.00636)	(0.00536)	(0.00538)	(0.00463)
	Average accuracy for random sampling	0.747***	0.794***	0.856***	0.890***
		(0.0127)	(0.0109)	(0.00863)	(0.00605)
8	Average increase due to adaptive sampling	0.00531*	0.0244***	0.0528***	0.0551***
		(0.00318)	(0.00277)	(0.00230)	(0.00207)
	Average accuracy for random sampling	0.725***	0.780***	0.843***	0.876***
		(0.00590)	(0.00493)	(0.00363)	(0.00288)

Table A.1 gives regression results for the simulations in section 7.2, and specifically figure 4. The difference in accuracy between random and adaptive sampling is evaluated after 5, 10, 15 and 25 questions. The constant is the average predictive accuracy of random sampling, and the parameter on adaptive is the treatment effect of adaptive sampling. Standard errors are clustered at the simulated individual level and indicated in parentheses. .

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B Construction of the indices

We ask respondents questions about their preferences over management practices in four hypothetical scenarios. The introduction of the four scenarios is as follows:

1. A manager is looking to hire a new employee for the main operation of the firm he works at. When recruiting new employees:

2. A production worker comes to a manager asking for a pay rise, when dealing with this request:

3. A customer has bought something from the firm on credit, and comes to the manager saying he cannot pay before the deadline. When dealing with this customer:

4. An employee comes to the manager with a complaint about another employee. When dealing with such a complaint:

For each of these scenarios, the respondents are asked, variations of the following questions on a Likert scale. These are always interpreted as a higher number representing less flexibility:

1. It is more important to hire the right person than to follow a clear recruitment procedure.

2. It is more important to find a new employee who shows initiative than someone who will follow procedures and the manager's directions.

3. The manager's decision largely depends on the specific situation and should not be governed by clear strict rules and procedures.

4. The manager's current decision on hiring should be consistently applied in the future.

Item (1) is used to construct the procedure index, item (2) is used to construct the centralisation index, item (3) is used to construct the rules index, and item (4) is used to construct the consistency index.

These four indices are constructed using inverse-covariance weighted (precision-weighted) of the items and normalised to a distribution with a mean of zero and a variance of one.

<table>
<tr><th colspan="5">Enumerator info: Enumerator, please go through this twice. First for scenario 1, then scenario 2.</th></tr>
<tr><th colspan="5">Rules Survey - Introduction</th></tr>
<tr><td colspan="5">As the final part of this survey we would like to ask experienced Ethiopian managers like you about their philosophy to managing. To do so, we would like to present to you some artificial scenarios to ask how you would act in these. Clearly, there are no right or wrong answers here. Based on what you tell us we will try to infer some patterns which may be consistent with your behaviour. These may also be interesting for you to examine your own decisions.
For each of these scenarios, we would like you to imagine I am a colleague coming up to you, sharing some information about a situation in the firm and asking you for your decision. Given the information I give you, you can also ask me for additional information. The goal is to both understand what information you find important making your decision, and what you actually decide.
For example, I may come to you with an employee who has missed a day of work without good reason. I would like to know whether we should fire this employee, but before answering you can ask whether the employee has had a previous and/or whether you think it is a good employee. You may then ask me about either of these facts, and then decide whether you would fire the employee.
Do you understand the protocol or have any questions?</td></tr>
<tr><th></th><th colspan="4">Scenario 1 and 2 - Current</th></tr>
<tr><td>s1</td><td colspan="2">Firm ID</td><td></td><td>|_____|</td></tr>
<tr><td>s2</td><td colspan="2">Scenario</td><td>01 = scenario 1
02 = scenario 2</td><td>|_____|</td></tr>
<tr><td>s3_1</td><td colspan="2">If Scenario 1:
Do you have influence on employee compensation at your firm?</td><td>01 = Yes
02 = No</td><td>|____|</td></tr>
<tr><td>s3_2</td><td colspan="2">If Scenario 2:
Do you have influence on employee hiring at your firm?</td><td>01 = Yes
02 = No</td><td>|____|</td></tr>
<tr><td colspan="5">Enumerator; tell the respondent:
Scenario 1: For our first set of scenarios we will talk about an employee who works in your main operation (e.g. an accountant in an accounting firm, a production worker in a manufacturing firm).
Scenario 2: For our second set of scenarios we will talk about hiring an employee for an entry-level position in your main operation.</td></tr>
<tr><td>s5</td><td>What job does an employee in your main operation do? [Enumerator; just record the job you are focussing on here if there are multiple]</td><td>01 = Construction
02 = Administration and Management
03 = Admin/Clerical/Office
04 = Accountant/ Finance/Purchaser
05 = Nurse/Health/Medicine
06 = Teacher/Tutor
07 = Lawyer
08 = Engineer/Architect
09 =Journalist
10 = Psychologist
11 = Banker
12 = Hotel Work (incl. waiter)</td><td>15 = Mechanic
16 = Machine Operator
17 = Businessman
18 = Trader/Sales/Retail
19 = Electrician
20 = Driver
21 = Statistician/ Data Collection
22 = Beauty/Hair/Salon 23 = Cleaner/Housework
24 = Transport/Taxi Work
25 = Cook/bakery
26 = Security/Guard/Soldier
27 = Entertainment/Art
28 = Church/Priest
29 = Other</td><td>|___|</td></tr>
</table>

<table>
<tr><td></td><td></td><td>13 = Factory
14 = Woodwork and Metal Work/Carpenter/Crafts</td><td></td><td></td></tr>
<tr><td>s5_oth</td><td>Specify other</td><td>If s5 = 29</td><td></td><td>|______|</td></tr>
<tr><td colspan="5">Enumerator, read out: In order to understand your responses better, we would like to limit the scope of information we share with you to the following set of characteristics.
Enumerator; share the sheet with the set of characteristics with the respondent and go through these. Order will be random on your sheets.</td></tr>
<tr><td colspan="5">Enumerator, tell the respondent:
Only for scenario 1: Suppose I, as your colleague, tell you an employee is unhappy with their pay and would like a raise [next point pay scale] and I want your decision.
Only for scenario 2: Suppose I, as your colleague, tell you I have a candidate for an entry-level position at your firm for which you are hiring. I would like your judgement of whether you would hire this candidate based on what I tell you.

Both scenarios: I will share some information with you, after which you can ask for as many additional facts from this list as you need to make your decision. We would like you to see these scenarios as entirely unrelated. For each scenario, you should decide whether you would give the pay rise/hire the employee, and what information you would like us to share.

Before we start, we would therefore like to ask you whether there is any information in this list that you would always ask you colleague for, regardless of what other information you already have. Secondly, we would like to ask the whether there are any facts you would never find relevant to your decision and thus never ask your colleague about. Answering honestly here will make the survey more relevant for you.</td></tr>
<tr><td>s6</td><td>Is there any information from this list you would always need to know to make your decision, regardless of what other information you have?</td><td>Enter the up to four characteristics.</td><td>C1
C2
C3
C4
C5
C6
C7</td><td>|__|
|__|
|__|
|__|</td></tr>
<tr><td>s7</td><td>Is there any information from this list you would never ask to know to make your decision.</td><td>Enter up to two characteristics.</td><td>C1
C2
C3
C4
C5
C6
C7</td><td>|__|
|__|</td></tr>
<tr><td colspan="5">Enumerator: After 7, press the done button and check the details on the pop-up window. If these are correct, press confirm to go to the first scenario. If they are not, change these and press done again before pressing confirm to start the first scenario.</td></tr>
<tr><td colspan="5">Repeat for 8 scenarios:
Share initial information
Ask for additional information wanted</td></tr>
</table>

- Record decision
- Add comment
- Press confirm

<table>
<tr><th></th><th></th><th></th><th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th></tr>
<tr><td rowspan="7">s8</td><td rowspan="7">What information would you like to have beyond what I told you [Enumerator; make sure the respondent only learns one characteristic at a time].</td><td rowspan="7">01 = Yes
02 = No

Only enter if 01</td><td>C1</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td></tr>
<tr><td>C2</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td></tr>
<tr><td>C3</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td></tr>
<tr><td>C4</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td></tr>
<tr><td>C5</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td></tr>
<tr><td>C6</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td></tr>
<tr><td>C7</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td></tr>
<tr><td>s9</td><td>What decision does the respondent make?</td><td>01 = Yes
02 = No</td><td></td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td></tr>
<tr><td>s10</td><td>Add any comments.</td><td></td><td>1</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td></tr>
</table>

Enumerator: Ask about the specific scenarios given by the program once these pop up. Here the respondent is not allowed to ask about additional information.

<table>
<tr><th></th><th>Number</th><th></th><th>1</th><th>2</th><th>3</th></tr>
<tr><td>S9_ext
Decision</td><td>What decision does the respondent make?</td><td>01 = Yes
02 = No</td><td>|______|</td><td>|_______|</td><td>|_______|</td></tr>
<tr><td>S10_ext
Comment</td><td>Add any comments.</td><td></td><td>|______|</td><td>|_______|</td><td>|_______|</td></tr>
</table>

Enumerator; when next window pops up:
We are now done with the first part of this scenario. Based on this, we have inferred some rules that other managers may use to make the same decisions as you in the scenarios we asked about. We would like to ask whether you think these patterns are consistent the decision making of your firm.

These rules are of the form: If I know a certain set of characteristics, I do not need any additional information. Going back to our first example before these vignettes a rule may look something like this: If an employee misses a day of work, has a previous warning for this behaviour and is bad at your job, no other characteristic of the employee will make you change your mind.

The rules we suggest here may apply to your decision making with what you told us so far, but because we have not asked about all combinations of the characteristics some may not apply to you.

Enumerator: The first rule will come up on the new window. Ask about this rule and the follow-up questions if the rule applies. Then press OK to go to the next rule. This part is done when the window closes.

<table>
<tr><th></th><th></th><th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th></tr>
<tr><td>s11</td><td>Does this rule apply</td><td>01 = yes
02 = No [Press confirm to skip to next rule]</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td></tr>
<tr><td>s12</td><td>What type of</td><td>01 = Formal</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td><td>|_|</td></tr>
</table>

	rule is this? Is this a formal, informal or personal rule	02 = Informal 03 = Personal Note: **Formal rule**: Written down in organisational documents **Informal rule**: Understood to apply by those in the organisation but not formalised **Personal rule**: Applied by this manager										
s13	Would your inferiors know this rule applies in the firm without asking you or their manager?	01 = Yes 02 = No	|_|	|_|	|_|	|_|	|_|	|_|	|_|	|_|	|_|	|_|
s14	Would you always be willing to apply this rule without asking your superior?	01 = Yes 02 = No	|_|	|_|	|_|	|_|	|_|	|_|	|_|	|_|	|_|	|_|

Enumerator info: Enumerator, please go through this twice. First for scenario 1, then scenario 2.				
Rules Survey - Training				
Enumerator, read out to respondent: As the final part of this survey we would like to ask experienced Ethiopian managers like you about their philosophy to managing. To do so, we would like to present to you some artificial scenarios to ask how you would act in these. Clearly, there are no right or wrong answers here. For each of these scenarios, we would like you to imagine I am a colleague coming up to you, sharing information about a situation in the firm and asking you for your decision. The goal is to understand how you would decide in these various scenarios. As an example, I may come to you asking you about an employee who has recently started at the firm who is quite good at their job, but whom has missed a day of work without good reason. I would like to know whether we should fire this employee. Obviously this is somewhat artificial, but we hope to understand your philosophy to these decisions. Do you understand the protocol or have any questions?`				
s1	Firm ID			|____|
s2	Scenario		01 = sc1 02 = sc2	|____|
s3_1	If Scenario 1: Do you have influence on employee compensation at your firm? [if no; skip]		01 = Yes 02 = No	|____|
s3_2	If Scenario 2: Do you have influence on employee hiring at your firm?		01 = Yes 02 = No	|____|
Enumerator; tell the respondent: **Scenario 1:** For our first set of scenarios we will talk about an employee who works in your main operation (e.g. an accountant in an accounting firm, a production worker in a manufacturing firm). **Scenario 2:** For our second set of scenarios we will talk about hiring an employee for an entry-level position in your main operation.				
s5	What job does an employee in your main operation do? [Enumerator; just record the job you are focussing on here if there	01 = Construction 02 = Administration and Management 03 = Admin/Clerical/Office 04 = Accountant/ Finance/Purchaser 05 = Nurse/Health/Medicine	15 = Mechanic 16 = Machine Operator 17 = Businessman 18 = Trader/Sales/Retail 19 = Electrician 20 = Driver 21 = Statistician/ Data	|___|

	are multiple]	06 = Teacher/Tutor 07 = Lawyer 08 = Engineer/Architect 09 =Journalist 10 = Psychologist 11 = Banker 12 = Hotel Work (incl. waiter) 13 = Factory 14 = Woodwork and Metal Work/Carpenter/Crafts	Collection 22 = Beauty/Hair/Salon 23 = Cleaner/Housework 24 = Transport/Taxi Work 25 = Cook/bakery 26 = Security/Guard/Soldier 27 = Entertainment/Art 28 = Church/Priest 29 = Other	

Enumerator, read out: In order to understand your responses better, we would like to limit the scope of information we share with you to the following set of characteristics.
Enumerator; share the sheet with the set of characteristics with the respondent and go through these. Order will be random on your sheets.

Enumerator, tell the respondent:
Scenario 1: Suppose I, as your colleague, tell you an employee is unhappy with their pay and would like a raise [next point pay scale] and I want your decision.
Scenario 2: Suppose I, as your colleague, tell you I have a candidate for an entry-level position at your firm for which you are hiring. I would like your judgement of whether you would hire this candidate based on what I tell you.
Both scenarios: I will share the information for a number of similar cases with you in which an employee asks for a pay rise. We would like you to see these cases as entirely unrelated. For each case, you only need to decide whether you would give the pay rise/hire the employee.

<table>
<tr><td></td><td></td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr>
<tr><td>s9</td><td>Decision
01 = Yes
02 = No</td><td>|____|</td><td>|_____|</td><td>|_____|</td><td>|_____|</td><td>|_____|</td><td>|_____|</td></tr>
<tr><td>s10</td><td>Comment</td><td>|_____|</td><td>|_____|</td><td>|_____|</td><td>|_____|</td><td>|_____|</td><td>|_____|</td></tr>
</table>

** Characteristics for Mail Merge

Scenario 1: Giving an employee a pay rise

C1: Performance compared to other employees with similar jobs (1=above average, 0=below average)

C2: Your colleague personally likes the employee (1=yes, 0 = no)

C3: The employee could easily leave for another firm (1=yes, 0 = no)

C4: The employee has had a pay rise in the past year (1=yes, 0 = no)

C5: Employee's pay is above average at the firm (1=yes, 0 = no)

C6: The employee has worked at the firm for more than two years (1=yes, 0 = no)

C7: The employee's gender (1=male, 0=female)

Scenario 2: Hiring a new employee

C1: The candidate showed enthusiasm for the job (1=yes, 0 = no)

C2: The candidate performed well in the interview (1=yes, 0 = no)

C3: The candidate speaks English (1=yes, 0 = no)

C4: The candidate fits with your firm’s culture (1=yes, 0 = no)

C5: The candidate has relevant experience at another firm (1=yes, 0 = no)

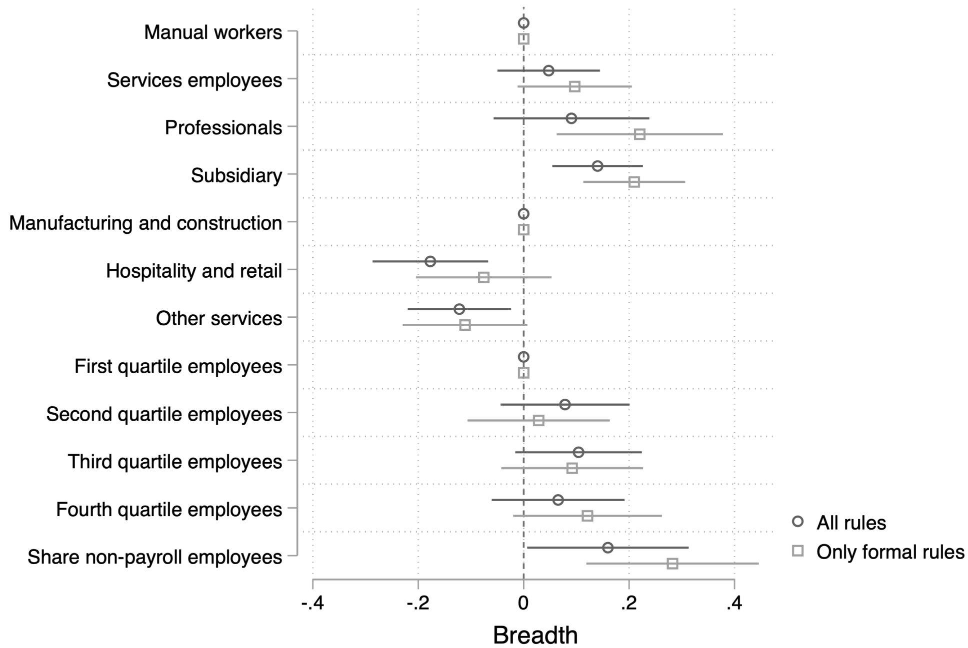
C6: The candidate has a university degree (1=yes, 0 = no)

C7: The candidate's gender (1=male, 0 = female)

C Additional results

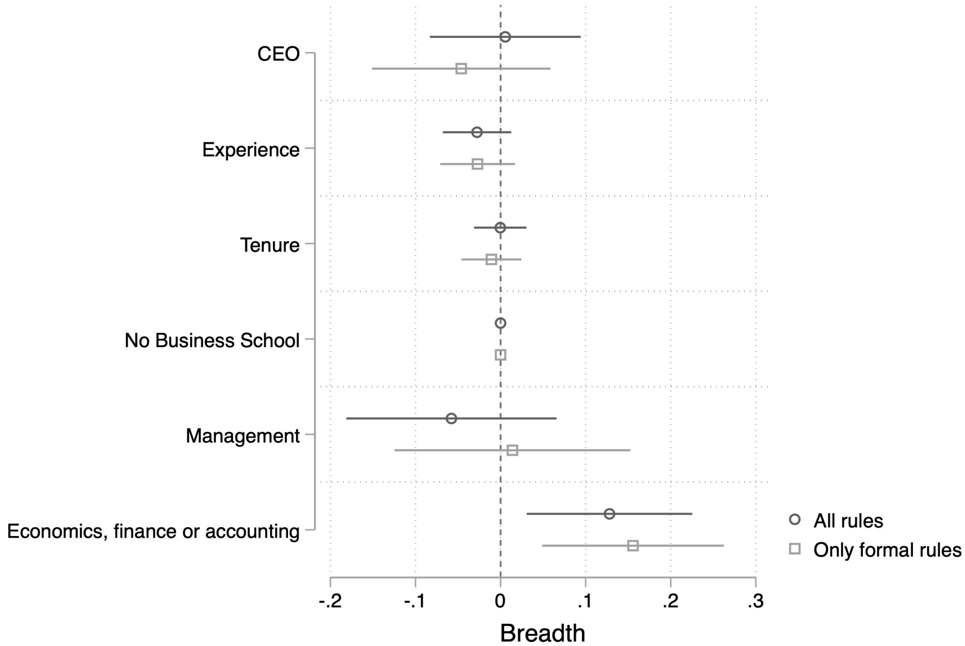
C.1 Firm and manager observables: Univariate regressions

Figure A.2: The relationship between firm observables and the breadth of the rulebook.



Notes: This figure reports regression results from univariate regressions of the breadth of the full and formal rulebook on various firm characteristics. These include employee type, categorized into manual workers, service employees, and professionals, with manual workers as the base category; the share of non-payroll employees; a subsidiary dummy variable where 1 denotes that the plant where the interview took place is a subsidiary of a larger entity; industry classification, which includes manufacturing and construction, retail and trade, and other service sectors such as tourism and education, with manufacturing as the base category; and firm size in terms of employment, included by quartile. Error bars represent 95% confidence intervals, with standard errors clustered at the firm level.

Figure A.3: The relationship between manager characteristics and the breadth of the rulebook.



Notes: This figure reports regression results from both univariate regressions of the breadth of the full and formal rulebook on various respondent characteristics. The CEO dummy indicates whether the respondent is the CEO or equivalent. Experience and tenure are respectively years of experience and years of experience at the firm, measured in units of five years. The management dummy indicates whether the respondent studied management at a business school. The economics dummy indicates if the manager studied economics, finance, or accounting at a business school. Error bars represent 95% confidence intervals.

C.2 Organisational Rules and Firm Performance

Table A.2: Organisational Rules and Firm Performance

	(1) Sales b/se	(2) Sales b/se	(3) Sales b/se	(4) Sales b/se
Panel A: Sales				
Management Score		-19.571		-22.056
		(18.68)		(19.22)
Breadth formal			64.524	70.661*
			(39.25)	(38.97)
Constant	-2.456	-4.046	-11.276	-20.655
	(33.07)	(33.42)	(25.38)	(33.06)
Controls	Yes	Yes	Yes	Yes
N	242	242	242	242
R2	0.1880	0.1979	0.2018	0.2149
Adj-R2	0.1672	0.1739	0.1849	0.1880

	(1) Profit b/se	(2) Profit b/se	(3) Profit b/se	(4) Profit b/se
Panel B: Profit				
Management Score		-0.577		-0.768
		(0.80)		(0.91)
Breadth formal			6.538**	6.632**
			(2.78)	(2.85)
Constant	1.017	0.908	0.164	-0.394
	(3.07)	(3.10)	(2.17)	(3.17)
Controls	Yes	Yes	Yes	Yes
N	232	232	232	232
R2	0.0592	0.0610	0.0898	0.0955
Adj-R2	0.0341	0.0316	0.0697	0.0631

	(1) Profit/Sales b/se	(2) Profit/Sales b/se	(3) Profit/Sales b/se	(4) Profit/Sales b/se
Panel C: Profit over sales				
Management Score		-0.006		-0.008
		(0.02)		(0.02)
Breadth formal			0.081*	0.083*
			(0.04)	(0.04)
Constant	0.203***	0.202***	0.186***	0.185***
	(0.05)	(0.05)	(0.05)	(0.05)
Controls	Yes	Yes	Yes	Yes
N	230	230	230	230
R2	0.0669	0.0675	0.0864	0.0877
Adj-R2	0.0418	0.0381	0.0576	0.0546

Notes This table examines how predictive the formal rule breadth is for sales, profit and profit over sales relative to the MOPS management score including the HR question. The calculation of the management score and formal breadth are detailed earlier in this section. All standard errors are clustered at the firm level. The included controls are the number of employees, sector, the type of position, management practices, and whether the firm is a subsidiary. Figure 1 details these variables, the number of employees is incorporated as a continuous variable rather than by quartile. Standard errors, clustered at the firm-level, are indicated in braces and significance is indicated as: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Details of the trained Gaussian Process

This section provides a more detailed description of the implemented Gaussian Process, the estimated parameters based on training data and their interpretation.

The linear mean function is represented as:

$$m(\mathbf{x}) = \mathbf{x}^\top \boldsymbol{\beta} + \alpha$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_7]^\top$ is the vector of characteristics (e.g., performance, tenure, etc.). $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_7]^\top$ are the coefficients representing the contribution of each characteristic to the mean function. α is the intercept term.

The covariance structure of the Gaussian Process is defined by a squared exponential kernel with Automatic Relevance Determination (ARD):

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{7}\frac{(x_{i,d} - x_{j,d})^2}{\ell_d^2}\right)$$

where σ_f^2 is the variance parameter, controlling the overall vertical variation, $x_{i,d}$ and $x_{j,d}$ are the d-th characteristics of the inputs $\mathbf{x}_i$ and $\mathbf{x}_j$, respectively. ℓ_d is the length-scale parameter for the d-th characteristic, controlling how much that characteristic influences the similarity between two input points (i.e., how sensitive the covariance is to changes in that particular characteristic).

The Gaussian Process prior over the latent function $f(\mathbf{x})$ is:

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}_i, \mathbf{x}_j)\right)$$

where $m(\mathbf{x})$ is the linear mean function and $k(\mathbf{x}_i, \mathbf{x}_j)$ is the squared exponential kernel with ARD.

The logistic link function used to map the latent output into the probability space is:

$$p(y = 1 \mid \mathbf{x}) = \sigma(f(\mathbf{x})) = \frac{1}{1 + \exp(-f(\mathbf{x}))}$$

where $f(\mathbf{x})$ is the latent output from the Gaussian Process and $\sigma(\cdot)$ is the logistic (sigmoid) function, which transforms the latent variable into the range $(0, 1)$.

These equations describe the components of your Gaussian Process model, including the linear mean function, the squared exponential kernel with ARD, and the logistic link function for probabilistic interpretation.

This table shows that the probability of a positive response is increasing in each of the characteristics. This is as expected for all the characteristics, except for gender where the relationship with decisions was ex-ante unclear. In future analysis, this could be implemented in a regression framework to better understand the results.

Table A.3: Parameters for Gaussian Processes: Pay Rise and Hiring Scenarios

Characteristic	Mean parameter	**Value**	Kernel parameter	**Value**
Performance compared to others	β_1	0.1404	ℓ_1	0.5132
Colleague likes the employee	β_2	0.2101	ℓ_2	0.5362
Employee could leave for another firm	β_3	0.2662	ℓ_3	0.9306
Had a pay rise in the past year	β_4	0.1667	ℓ_4	0.0451
Pay is above average at firm	β_5	0.1038	ℓ_5	0.5377
Worked at firm > 2 years	β_6	0.1821	ℓ_6	1.3443
Employee's gender, male=1	β_7	0.1623	ℓ_7	0.2442
	α	-0.8449	σ_f^2	0.3197
Characteristic	Mean parameter	**Value**	Kernel parameter	**Value**
Candidate showed enthusiasm	β_1	0.2092	ℓ_1	0.6234
Performed well in interview	β_2	0.2943	ℓ_2	0.0914
Speaks English	β_3	0.4317	ℓ_3	0.8466
Fits firm's culture	β_4	0.3491	ℓ_4	0.5290
Relevant experience	β_5	0.2273	ℓ_5	1.6567
Has university degree	β_6	0.2545	ℓ_6	0.3127
Candidate's gender, male=1	β_7	0.2252	ℓ_7	0.3207
	α	-1.2087	σ_f^2	0.1289

In the pay-rise scenario, the probability of the decision being one varies from 30% when $X = 0$ for all characteristics to 60% when $X = 1$ for all characteristics. For the hiring scenario, this range is broader and goes from 23% to 69%.

References

Abebe, G., M. Fafchamps, M. Koelle, and S. Quinn (2019). Learning management through matching: A field experiment using mechanism design. Technical report, National Bureau of Economic Research.

Abebe, G., M. Fafchamps, M. Koelle, S. Quinn, and T. Schwantje (2025, January). Management style under the spotlight: Evidence from studio recordings. Cepr discussion paper, Centre for Economic Policy Research, London.

Adler, P. S. and B. Borys (1996). Two types of bureaucracy: Enabling and coercive. *Administrative Science Quarterly 41*(1), 61–89.

Aghion, P., N. Bloom, B. Lucking, R. Sadun, and J. Van Reenen (2021). Turbulence, firm decentralization, and growth in bad times. *American Economic Journal: Applied Economics 13*(1), 133–169.

Aghion, P. and J. Tirole (1997). Formal and real authority in organizations. *Journal of Political Economy 105*(1), 1–29.

Akcigit, U., H. Alp, and M. Peters (2021a). Lack of selection and limits to delegation: firm dynamics in developing countries. *American Economic Review 111*(1), 231–275.

Akcigit, U., H. Alp, and M. Peters (2021b, January). Lack of selection and limits to delegation: Firm dynamics in developing countries. *American Economic Review 111*(1), 231–75.

Atkin, D., A. K. Khandelwal, and A. Osman (2019). Measuring productivity: Lessons from tailored surveys and productivity benchmarking. In *AEA Papers and Proceedings*, Volume 109, pp. 444–449. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Bandiera, O., A. Prat, S. Hansen, and R. Sadun (2020). Ceo behavior and firm performance. *Journal of Political Economy 128*(4), 1325–1369.

Bassi, V., J. H. Lee, A. Peter, T. Porzio, R. Sen, and E. Tugume (2023). Self-employment within the firm. Technical report, National Bureau of Economic Research.

Bloom, N., E. Brynjolfsson, L. Foster, R. Jarmin, M. Patnaik, I. Saporta-Eksten, and J. Van Reenen (2019). What drives differences in management practices? *American Economic Review 109*(5), 1648–1683.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013). Does management matter? evidence from india. *The Quarterly journal of economics 128*(1), 1–51.

Bloom, N., C. Genakos, R. Sadun, and J. Van Reenen (2012). Management practices across firms and countries. *Academy of management perspectives 26*(1), 12–33.

Bloom, N., A. Mahajan, D. McKenzie, and J. Roberts (2010). Why do firms in developing countries have low productivity? *American Economic Review 100*(2), 619–623.

Bloom, N., R. Sadun, and J. Van Reenen (2016, May). Management as a technology? Working Paper 22327, National Bureau of Economic Research.

Bloom, N. and J. Van Reenen (2007). Measuring and explaining management practices across firms and countries. *The quarterly journal of Economics 122*(4), 1351–1408.

Callander, S. and N. Matouschek (2019). The risk of failure: Trial and error learning and long-run performance. *American Economic Journal: Microeconomics 11*(1), 44–78.

Caria, A. S., G. Gordon, M. Kasy, S. Quinn, S. O. Shami, and A. Teytelboym (2024). An adaptive targeted field experiment: Job search assistance for refugees in jordan. *Journal of the European Economic Association 22*(2), 781–836.

Chapman, J., E. Snowberg, S. W. Wang, and C. Camerer (2024). Looming large or seeming small? attitudes towards losses in a representative sample. *Review of Economic Studies*, rdae093.

Englmaier, F., J. E. Galdon-Sanchez, R. Gil, M. Kaiser, and H. Strandt (2025). Management practices and firm performance during the great recession. Technical report, CRC TRR 190 Rationality and Competition.

Frazier, P. (2018). A tutorial on bayesian optimization. *ArXiv abs/1807.02811*.

Gibbons, R. and R. Henderson (2012). Relational contracts and organizational capabilities. *Organization Science 23*(5), 1350–1364.

Ichniowski, C., K. Shaw, and G. Prennushi (1997). The effects of human resource management practices on productivity: A study of steel finishing lines. *The American Economic Review 87*(3), 291–313.

Kasy, M. and A. Sautmann (2021). Adaptive treatment assignment in experiments for policy choice. *Econometrica 89*(1), 113–132.

Li, D., L. R. Raymond, and P. Bergman (2020, August). Hiring as exploration. Working Paper 27736, National Bureau of Economic Research.

Li, J., A. Mukherjee, and L. Vasconcelos (2022, September). What makes agility fragile? a dynamic theory of organizational rigidity. Article 69(6), Management Science.

March, J. G. and H. A. Simon (1958). *Organizations*. New York: Wiley.

Press, I. P. (1989). Taiichi ohno, toyota production system (beyond. *Human Systems Management 8*, 175–182.

Syverson, C. (2011). What determines productivity? *Journal of Economic literature 49*(2), 326–365.

Taylor, F. W. (1911). Principles and methods of scientific management. *Journal of Accountancy 12*(3), 3.