

Management Style Under the Spotlight: Evidence from Studio Recordings

Online Appendix

February 2026

Contents

A	Vignette scripts	4
A.1	Respondent management vignettes	4
A.2	Assessment of management styles: Hypothetical workplace conflict	6
B	Experiment	11
B.1	Structure of the assessments	11
B.2	Matching firms to candidates	12
B.3	A revealed preference exercise with human resource managers	12
B.4	A stated preference exercise with human resource managers	13
C	Details: Encoding of Responses	14
D	Further detail on intentions	16
E	Bayesian modelling in Stan	18
E.1	Dirichlet model	18
E.2	Plackett-Luce model	20
E.3	Estimation	23
E.3.1	The Dirichlet model	23
E.3.2	The Plackett-Luce model and joint-estimation	23
E.3.3	Stability of type parameters with joint estimation	26
F	Sensitivity to the choice of number of types	29
F.1	Considering two types and three types	29
F.1.1	Results: Two-type model	31
F.1.2	Results: Three-type model	38
F.2	Considering five types	45
G	Inclusion in the studio	46
H	Variations on the Plackett-Luce model	49
H.1	The distribution of β_f conditional on covariates	49
H.2	Allowing firm preferences to vary by vignette	49
H.3	Allowing firm preferences to vary by candidate gender	50
H.4	Overcontrolling for the duration of the response	50
H.5	Causal effect of managerial experience	51
I	Describing heterogeneity in firm preferences	51
J	Worker preferences over managers	54
K	Summary statistics	62

L	Robustness Gender Tables	64
M	Text analysis	69
M.1	Text cleaning	69
M.2	Word clouds	70
M.3	Text features predicting types	73
M.4	Semantic word clusters and managerial styles	81
M.5	Clusters with the strongest positive correlation with each type	83
M.6	Person-Level Text Classifier	85
N	AI Analysis	87
N.1	Results from the AI Survey	87
N.2	What words describe responses?	94

A Vignette scripts

A.1 Respondent management vignettes

Vignette 1: An absent employee

NARRATOR: *In this video, an accounting clerk who has been absent for three days without advance notice to her immediate boss is shown. (S)he shows up for work on the fourth day, explaining that (s)he was unwell. Do watch the video carefully.*

THE CLERK: I realize that I did not come to work for the last three days. I could not also give you any advance notice as this was something beyond my control. I was not feeling well. I had a bad flu that made me see a doctor. The doctor recommended that I should take it easy and perhaps take a few days of rest at home. I thought that, with my conditions, coming to the office and working on our accounts would be difficult [actor coughs slightly]. Also, I did not want others to catch my flu as well, so I stayed away.

Boss, you remember Almaz, my colleague, who had a bad flu three weeks ago. Instead of staying at home, she chose to come to work while being very sick. The next day, almost half of the staff in our department had a terrible headache and were coughing. They caught whatever she had. I certainly did not want that to happen [actor shakes head in the negative].

Vignette 2: Negotiating with a supplier

NARRATOR: *Here we have a supplier, who explains to you that (s)he cannot fulfil an order according to specifications, because of problems sourcing input materials. (S)he offers to supply a replacement of inferior quality instead. Note that quality is the most important feature of the firm's product and that the firm is known in the market for producing and selling the highest quality products in the industry. Do watch the video carefully.*

THE SUPPLIER: As you know, the main roads that we use to transport our shipments from Djibouti have been blockaded in the past several weeks. It has been three weeks since we got our last shipment of the materials that we promised to supply to you. During the course of the past three weeks, our inventory of these materials has been depleted. I am thus sorry to inform you that we cannot supply you with the materials as specified in the original supply agreement.

However, the good news is that we have a pretty good stock of alternative materials that we could supply you at a cheaper rate. We cannot, however, guarantee that these materials currently in stock meet the quality and specification requirements that were stipulated in the original supply agreement. Nevertheless, I think what we have for you is a good deal that is offered with the cheapest possible price. Of course, once our sourcing problem is resolved, we will continue to supply you with materials that comply with the specifications that you set.

Vignette 3: An employee seeks a pay rise

NARRATOR: *In this video, we will show you a production worker, coming to you and asking for a pay rise. Your company does not have a plan to increase any of your workers' salary this year. Do watch the video carefully.*

THE PRODUCTION WORKER: Boss, you know how dedicated I am to my work. In the past three months, our department has done very well and we have met the quarterly targets set by the management. I'm very happy about that. As you know, I have been the most productive worker in the department. I am one of the few people in the department that exceeded the personal productivity targets that you set for us at the beginning of the year. Of course, I also did not have even a single day of absence recorded in the completed quarter. If you don't believe me, you can check the attendance repository and the productivity (or KPI) sheet. As you know, I also worked for the company for the past ten years, not so many people can say that! I am not sure whether you know my mother, Feleku Kuma; she is widely known around here, she also worked here for a very long time.

I think these speak volumes to my dedication and loyalty to this company. But honestly, I don't think my current pay reflects the effort that I exert and the loyalty that I and my family have for the company. In fact, I spoke to people who work with our competitors at the same type of occupation as I am; surprisingly, their pay is much larger than mine. I am sure you agree with me that I deserve a pay rise. Can you let me know as to when and how much I should expect to get as a pay rise?

Vignette 4: Negotiating an adjustment with the bank

NARRATOR: *This video shows a bank manager, who calls to remind you about an unpaid loan installment. But you will not have sufficient funds to pay the bank for another two weeks. Take a watch.*

THE BANK MANAGER: Thanks a lot for taking my call. As you know, you are one of our clients who missed the latest installment on your outstanding loan from our bank. Our bank treats its customers with utmost respect and encourages timely payment to avoid breaches to loan terms and conditions. But also, we have quite a lot of pressure from the board of directors to ensure that scheduled loan payments are made on time. Failure to pay loan installments not only increases the interest payment you owe us; it also reduces your credit rating with us. Next time you ask for loan, you may be denied access altogether – or may be eligible to only a fraction of the amount that you have applied for. I thus cannot stress enough the importance of settling your outstanding installment as soon as you can. That said, I am wondering whether you will be able to pay the unpaid loan installment very soon. I will follow-up this call with a written reminder to alert you to the periodic payment that we expect to be paid from your company.

Vignette 5: Negotiating with a client

NARRATOR: *The next video shows a client of your business, who has not paid what you have invoiced, in spite of one reminder. The client comes to place a new order.*

THE CLIENT: I really like the service I got from your company last time. I don't think I had the chance to thank you for that – so thank you. Oh, I did not forget about the payment and I confirm that I have received the invoice you issued. But you know how it is these days, business is not great, and our product is not exactly flying off the shelves. We are really short on cash at the moment. I will thus have to ask you to bear with me a little to pay for the previous order. With the risk of stretching your generosity, I'm here today to place another order. In a spirit of our long-standing relationships, I appreciate if you could deliver my new order soon without waiting for payment for the previous order. I promise we will meet our payment obligations for the previous and new orders once we sort out our cash flow problem.

A.2 Assessment of management styles: Hypothetical workplace conflict

Setting up the scenario

NARRATOR: You are faced with a situation where workers go on strike demanding proper latrines. A representative of the workers comes to talk to you. Take a look at the video carefully.

THE WORKER REPRESENTATIVE: In my capacity as a representative of workers in this factory, I would like to speak to you about the ongoing strike in the factory. I am sure that by now you realize what workers are striking about. But with the slow response from the line supervisors and the deafening silence from you, I am not certain that the management in the factory understood the severity of the problem. That is why I'm here.

Look, the strike is about the poor toilet facilities in the factory. As you recall, during the worker/management meeting, many workers complained about the lack of sufficient toilet facilities; there are very few toilets in the compound and workers have to walk to the other end of the compound to access them. Even when they reach the toilets, they have to wait their turn, sometime for several minutes, to use them. In recent months, the problem has gotten worse. The few toilets that are available either do not have water or the flushing system does not work and tissue papers are not available. To make things worse, sometimes the toilets are locked and workers are not told why they are locked and who is keeping the keys. Further, even when workers try to use the toilets, they face restrictions on the number of toilet breaks they are entitled to take.

The attitude of your supervisors towards toilet use is also appalling. Our female workers suffer a great deal particularly when they are experiencing their period. Studies indicate that failing to use toilets on the course of a day lead to several health problems including damage to the bladder and several types of infections that have serious implications on the workers'

wellbeing. We as workers are very much aware of the implications of poor toilet facilities. Please do not underestimate our intelligence, we know about all this and our rights. Through our discussion with fellow workers, we came to the conclusion that our voice needs to be heard. We are particularly troubled by the dominant view of the management that workers shirk using toilet breaks as an excuse. Such views are not only unfounded, we reject the thought process behind it as it deprives the dignity of workers as human beings. Unless our demand for improved toilet facilities is not met, our strike will continue.

Six possible responses to the scenario

We then use different actors to record six separate possible managerial responses. The responses are designed to represent different styles of engagement. In this summary, we suggest several descriptors for each managerial response; these descriptors were not provided to the respondents.

NARRATOR: After listening to the complaint by the worker representative, six different managers respond in different ways. Please do watch the videos.¹

MANAGER 1 [*apologetic; collaborative; empowering*]: We understand workers' complaints about the poor state of toilet facilities in our factory. We admit that we have been slow to act and hence invited the strikes onto ourselves. On behalf of the company, I would like to apologize for our failure to nip this problem in its bud before it reached to this level. We will take immediate actions to rectify the situation. For a start, we will constitute a committee that will include workers' and management representatives. The committee will work on identifying the key problems in the existing toilet facilities and would suggest solutions for improving access and quality of toilet facilities. In parallel, the management will meet and earmark the necessary resources required to improve the facilities. We actually want the whole process to be led by workers and their representatives. We believe that there is no substitute for better working conditions to get the best out of our workers. With this pledge, we would like you to convince the workers to go back to work.

MANAGER 2 [*acknowledging; proactive; realistic*]: Honestly, this is a fair request. We should have seen this coming. As we promised during the worker/management meeting, the management takes all workers' requests very seriously. We are now making preparations to invest in improving the toilet facilities. Some of the improvements related with maintaining the flushing system of the existing facilities can be readily made. But increasing the number of toilets and digging septic tanks could take longer time. As you know, obtaining construction permit in this part of the city takes quite a long time. So even if we want, we may not be able to overhaul the whole facility at once. We thus ask for patience on the workers end. With this in mind and with the promise to make feasible changes immediately, we request you to stop the workers'

¹ For clarity, these are ordered based on the senior managers' ranking which was used to incentivise the respondents. Respondents' instead saw these in a random order.

strike at once.

MANAGER 3 [*appreciative; conciliatory; blame-shifting; behavioural focus*]: Thanks for bringing this to our attention before things go out of hand. We really appreciate your effort to be a voice of workers. The way you represent the concerns of your workers has been exemplary – and, on behalf of the company, I would like to express my gratitude. Of course, even if you are the representative of workers, we always felt that you are one of us, the management team. As both a worker representative and as someone who works closely with us, you understand how it has been difficult to manage workers, particularly their use of toilet facilities. We agree that facilities should improve, and that can be done.

But this issue is not that important. What worries us the most is workers have been abusing their toilet breaks by sitting idle and chattering with fellow workers during the breaks. Some of them spend more than 50 minutes inside the toilet, locked inside with their phone, checking Facebook or playing games. We told them repeatedly that this is not an acceptable behavior and it ought to change. When they refused, we started locking some of the toilet rooms. I think you should help us to get their act together. We can work together to change their behavior use their rightful toilet breaks only for intended purposes. Explain to them that the strike should come to an end and they should stop abusing their breaks. We can then discuss about ways to improve the quality of the facilities.

MANAGER 4 [*acknowledging; deflective; postponing; subtle pressure*]: We understand that toilet facilities are important and workers have the right to complain about problems in accessing these facilities in our compound. But everyone knows that we are struggling at the moment: business is not good and we are making losses. Asking the management to invest in improving the toilet facilities at the moment is not a good idea. This simply is not the right time to do that.

As a representative of workers, you should be aware that management does not like to be ambushed or cornered into making decisions in difficult times. I am sure you also do not want to take the blame for creating trouble in the company while conditions are already hard on us. I know you are also concerned about your career prospects. This is really a tough time for everyone, so let us weather the storm together and we can resume the discussion about improving the toilet facilities afterwards.

MANAGER 5 [*dismissive; threatening; defensive*]: As you said, we have discussed this during the worker-management meeting. During the meeting, we were very clear. We can only meet demands that are convincing and consistent with the company's policy. We told the workers that we know what kind of toilet facilities other employers provide for their workers.

Did you have a chance to visit other company's toilet facilities? If you or your colleagues had done that, you would have realized that ours is not of sub-standard quality compared to them. If anything at all, we provide flushing toilets and fresh water, which many don't. We also know the kind of toilet facilities workers have at their place of residence, mostly pit latrines

that are unsafe and often dirty. We certainly provide better facilities than those at home. So if you don't like our toilets here, you and your comrades can go and work elsewhere. And believe me, this is what is going to happen if you don't cease striking in the next 24 hours.

MANAGER 6 [*legalistic; hostile; threatening; dismissive*]: Who do you think you are, barging into my office and putting unwarranted demands on this company? I am not going to discuss about toilets or about any related issues with you. No, not at all. Forget toilets, let us talk about the strikes. Do you know what the labor law says about strikes? I am sure you don't! Had you known, you wouldn't be striking! Let me enlighten you. You see this [actor shows a copy of a labour proclamation]: this is Labour Proclamation No. 377/2003.

According to Article 158 (1) of the proclamation, before any strike happens, workers must give advance notice to the company explaining in detail the reasons for the strike. As far as I know, no such notices have been submitted to my office. Further, the law says that, for the strikes to be carried out, it should be supported by the majority of workers with a formal meeting. To my knowledge, no such formal meeting has happened. These two violations by you and your comrades indicate that the current strike is clearly in the contravention of the labor law. If the strikes does not stop before the end of the day, you will hear from our lawyers. We will sue you personally for initiating and implementing the strikes in defiance of the country's law.

Table A.1 summarises these responses.

Table A.1: Summary of Managerial Responses to Hypothetical Workplace Conflict

Ranking	Response
Manager 1	Apologetic, collaborative, empowering <ul style="list-style-type: none"> - Apologizes for slow response. - Proposes forming a committee with workers' involvement to address issue. - Pledges resources and seeks cooperation to end strike.
Manager 2	Acknowledging, proactive, realistic <ul style="list-style-type: none"> - Accepts request as fair and promises improvements. - Details immediate and long-term actions needed. - Requests patience and asks for strike to end.
Manager 3	Appreciative, conciliatory, blame-shifting, behavioral focus <ul style="list-style-type: none"> - Thanks representative and recognizes issue. - Blames workers for misusing toilet breaks. - Suggests ending strike and working together to improve behavior before addressing facilities.
Manager 4	Acknowledging, deflective, postponing, subtle pressure <ul style="list-style-type: none"> - Recognizes importance of issue but cites current financial struggles. - Suggests postponing improvements and implies potential career consequences for pushing issue.
Manager 5	Dismissive, threatening, defensive <ul style="list-style-type: none"> - Claims company's facilities are sufficient. - Compares to other companies' and workers' home facilities. - Threatens workers to cease strike or face consequences.
Manager 6	Legalistic, hostile, threatening, dismissive <ul style="list-style-type: none"> - Refuses to discuss issue. - Cites labour law violations and threatens legal action. - Demands strike to end immediately.

Notes This table summarises the managerial responses to the hypothetical workplace conflict, ordered based on the ranking of the senior manager (best to worst). The adjectives describing each of the vignettes were generated by ChatGPT.

B Experiment

B.1 Structure of the assessments

To implement the assessments, we first randomly divide our full set of candidates into groups of 15, and our full set of firm managers into groups of 10. We divide these sets of 15 candidates into five *triplets* (labeled 1 through 5) and 10 firms into five *pairs* (labeled A through E). As shown in Table A.2, each firm pair assesses different triplets of candidates in a rotating sequence across five vignettes:

- In the first vignette, pair A assesses triplet 1, pair B assesses triplet 2, pair C assesses triplet 3, and so on (see the diagonal in Table A.2).
- Keeping the same triplets and firm pairs, we rotate assignments in subsequent vignettes: for example, in the second vignette, pair B assesses triplet 1, pair C assesses triplet 2, pair D assesses triplet 3, etc.
- This pattern continues until each firm pair has assessed every candidate triplet exactly once.

We implement this procedure for 66 separate groups of 15 candidates, each assessed by a total of 66 groups of eight or ten firms. For the groups of eight firms, for each triplet of candidates one vignette is not assessed.

This rotating design enables us to measure how similarly two HR managers in the same firm pair rank the same candidates within a given vignette (i.e., *within-vignette* agreement) and how these rankings compare across vignettes (i.e., *cross-vignette* agreement). It thus provides a direct way to study consistency in managers' preferences.

Table A.2: Structure of the Assessments

	Firm Pair				
Candidate Triplet	A	B	C	D	E
1	V1	V2	V3	V4	V5
2	V5	V1	V2	V3	V4
3	V4	V5	V1	V2	V3
4	V3	V4	V5	V1	V2
5	V2	V3	V4	V5	V1

Notes: This table illustrates how one group of 15 candidates is assessed by one group of 10 firms (grouped into five pairs). Each cell shows the vignette (V1–V5) that a particular firm pair assesses for a particular candidate triplet.

B.2 Matching firms to candidates

We match firms to candidates using a random sequential dictator algorithm. The matching procedure is structured as follows:

- (i). We divide firms into groups of 10 and, separately, candidates into groups of 15.
- (ii). Each group of candidates is assigned a random preference ordering over all groups of firms.
- (iii). Groups of candidates are matched sequentially based on their preference rankings:
 - Each group of candidates is assigned to their most preferred group of firms that has not yet been matched.
- (iv). We impose an additional constraint: firms cannot assess candidates who have previously interned at their firm, as part of the experiment conducted in (Abebe, Fafchamps, Koelle, and Quinn, 2026).

Within each group-level match, triplets of candidates are matched to pairs of firms for the first round of assessments using the same algorithm. After this, triplets of candidates rotate across pairs of firms using the mechanism described in Section B.1.

B.3 A revealed preference exercise with human resource managers

We designed these mechanisms to achieve three objectives: (i) the mechanisms are intuitively simple for respondents to understand, and truthful reporting should be an ‘obviously dominant strategy’ (within the formal definition of Li (2017)); (ii) as far as possible, the two mechanisms are the same (so that any differences in reporting are attributable to differences in preferences, rather than to the mechanisms themselves), and (iii) we do not mislead the respondents in any way (therefore, any promises that we make in order to incentivise the decision are feasible for us to implement). Our mechanisms are as follows:

- (i). Suitability as an employed manager: We tell the respondent:

I’m about to ask the computer to randomly choose a number: 1 or 2. This is the number of candidates whose details I will be asked to pass to you. I need you to commit, in advance, to which candidates you would like to see in each circumstance. You have seen three candidates. Suppose that the computer tells me that you may receive the contact details of two. In that situation, of the three candidates, whose details would you then not want to see? Now, suppose that the computer tells me that you may receive the contact details of only one. Of the two candidates remaining, whose details would you then not want to see?”

This process continues until just one candidate remains.²

- (ii). Suitability as an entrepreneur: We explain to the respondent that we plan to run a series of business plan competitions – in which a set of candidates is judged by experienced business managers, with a winner receiving US\$1000 to start (or to support) his or her own business. To elicit a ranking over candidates’ entrepreneurial abilities, we nest the previous mechanism in a simple ‘random dictator’ mechanism. We tell the respondent:

You are one of two business people who will review this set of three aspiring managers. We will randomly choose just one of the two of you to decide which of the three should be candidates at the business plan competition. Suppose that person is you. The computer will, again, randomly choose a number: 1 or 2. This time, this number will tell us how many candidates will be invited to the business plan competition. You have seen three candidates. Suppose that you are chosen to send candidates to the business plan competition, and that the computer decides that you should send two candidates. In that situation, of the three candidates, who should not be sent to the competition?

As above, this process continues until just one candidate remains.

These mechanisms are each similar to the ‘OSP-RSD’ ranking mechanism described in Li (2017); they are simple to understand, and it is an ‘obviously dominant’ strategy for the respondent to rank truthfully.

B.4 A stated preference exercise with human resource managers

After each recording, we ask the manager the following questions (each of which uses a five-point Likert scale):

- (i). Suitability as an employed manager: *Think about the lowest-level managerial position in your firm. Imagine that a firm – either your firm or another firm like yours – is interested in hiring for this position. Based on the recording that you have just heard, how likely are you to recommend hiring this person for a managerial position like that?*
- (ii). Suitability as an entrepreneur: *Suppose that the person you have just seen is acting in the capacity as a founder and managing director of his/her own firm. In your opinion, how likely is it that this person would be a successful entrepreneur?*

These answers record managers’ stated preferences about each candidate separately.

² In a follow-up survey with the candidates, twelve report having been contacted by firms, none of these were eventually hired.

C Details: Encoding of Responses

Dimension	Enumerator encoding	Encoding for analysis
Action	1 Agrees	1 Agrees
	2 Disagrees	2 Disagrees
Justification	Varies by vignette; see following table	1 Firm's Interest
		2 Other person's interest
		3 Respondent's own interest
		4 Shared Interest
		5 No justification
Authority	1 Formal Authority	1 Formal Authority
	2 Higher Principles	2 Higher Principles
	3 Personal Authority	3 Personal Authority
	4 Personal Relationship	4 No Source
	5 No Source of Authority	
Tone		1 Aggressive
		2 Assertive
		3 Calm/Assured
		4 Timid

Table A.3: Encoding of Responses

Notes: This table reports the attributes encoded by our expert enumerators. In the analysis, we merge the Calm/Assured and Timid tone. We note that respondents virtually never refer to the personal relationship between themselves and the other party, and thus encode this as no source of authority.

Encoding Enumerator	Encoding for analysis
Line Management of an Employee	
The action is to protect/help the respondent's company	1. Firm's Interest
The action is in the interests of the other employees.	
The action is to protect/help the employee	2. Other party's interest
The action is in the respondent's own interest	3. Respondent's own interest
The respondent did not justify his or her action	5. No justification
Negotiating with a Supplier	
The action is to protect/help the respondent's company	1. Firm's Interest
The action is to protect/help the supplier	2. Other party's interest
The action is in the respondent's own interest	3. Respondent's own interest
Reference to the relationship between the respondent's firm and the supplier	4. Shared interest
The respondent did not justify his or her action	5. No justification
Other*	
Negotiating a Pay Rise	
The action is to protect/help the respondent's company	1. Firm's Interest
The action is to protect/help the employee	2. Other party's interest
The action is in the respondent's own interest	3. Respondent's own interest
Reference to the relation with the employee and his/her family	4. Shared interest
Reference to the employee's good performance	
Reference to the effect it might have on other employees	1. Firm's Interest
The respondent did not justify his or her action	5. No justification
Other*	5. No justification
Negotiate an Adjustment with the Banks	
The action is to protect/help the respondent's company	1. Firm's Interest
The action is to protect/help the bank	2. Other party's interest
The action is in the respondent's own interest	3. Respondent's own interest
Reference to the relationship between the firm and the bank	4. Shared interest
Reference to the inability of the company to repay the loan	1. Firm's Interest
The respondent did not justify his or her action	5. No Justification
Other*	5. No justification
Negotiate with a Client	
The action is to protect/help the respondent's company	1. Firm's Interest
The action is to protect/help the client	2. Other party's interest
The action is in the respondent's own interest	3. Respondent's own interest
Reference to the relationship between the respondent's firm and the client	4. Shared interest
The respondent did not justify his or her action	5. No justification
Other*	5. No justification

Table A.4: Encoding justification for our scenarios

D Further detail on intentions

Table A.5: **Reported intentions of types by vignette**

Vignette	Type	Expresses distrust counterparty	Maintain relationship	Follow procedure	Set an example
Rule-based	.559		.419	.572	
Affiliative	.429		.341	.504	
Power-based	.487		.376	.465	
Discretion-based	.444		.324	.475	
<i>p</i> -value	.003		.058	.058	
Rule-based	.031	.001	.283	.544	
Affiliative	.029	.006	.228	.447	
Power-based	.03	.002	.216	.459	
Discretion-based	.039	.002	.175	.435	
<i>p</i> -value	.664	.231	.006	.006	
Rule-based		.121	.59	.325	
Affiliative		.176	.371	.225	
Power-based		.08	.494	.198	
Discretion-based		.068	.476	.213	
<i>p</i> -value		0	0	0	
Rule-based	.002	.71			
Affiliative	0	.714			
Power-based	.001	.642			
Discretion-based	.004	.609			
<i>p</i> -value	.889	.007			
Rule-based	.606	.205	.439	.072	
Affiliative	.341	.434	.221	.017	
Power-based	.579	.207	.348	.028	
Discretion-based	.543	.194	.308	.018	
<i>p</i> -value	<0.001	<0.001	<0.001	<0.001	

Notes This table describes the intentions of respondents' for each vignette, which are enumerated before the respondent actually responds to the vignette. The four characteristics are (i) whether the respondent expresses they do not trust the other part, (ii) whether the respondent wishes to maintain a good relationship, (iii) whether the respon-

E Bayesian modelling in Stan

E.1 Dirichlet model

We observe individuals $i \in \{1, \dots, N\}$ performing on vignettes $v \in \{1, \dots, 5\}$. For each individual assessment of a vignette, we have a set of two enumerators, $e \in \{1, 2\}$. Each enumerator records a set of ‘attributes’ of the response (action, authority, justification and tone): $a \in \{1, \dots, 4\}$. Each attribute a has $J(a)$ possible categorical responses.³ $w_{ive}^a \in \{1, \dots, J(a)\}$ is the response recorded by enumerator e for vignette v for individual i for attribute a .

For attribute a , consider a ‘pure type’ $k \in \{1, \dots, K\}$, having a vector ϕ_{ka} of dimension $J(a)$, such that $\phi_{ka} \in (0, 1)^{J(a)}$ and $\sum_s \phi_{ka}^{(s)} = 1$. Denote by $\tilde{\phi}_{ka}$ the inverse Multinomial Logit transformation of ϕ_{ka} .⁴ For simplicity and tractability, we assume that, conditional on ϕ_{ka} , the attributes are realised independently as:

$$w_{ive}^a \mid \phi_{ka}, \psi_{av}, \chi_{ae} \sim_{iid} \text{Multinomial Logit} \left(\tilde{\phi}_{ka} + \psi_{av} + \chi_{ae} \right), \quad (1)$$

where ψ_{av} and χ_{ae} here represent $J(a)$ -element vectors.⁵⁶

This is essentially a Latent Dirichlet Allocation model, with a few adjustments for our setting:

- (i). ψ_{av} and χ_{ae} provide a measurement scheme; this is similar in spirit to Item Response Theory. (That is, this captures the notion that different enumerators will have different underlying tendencies to assess different attributes on different vignettes.) Note that, in the special case that $\psi_{av} = \mathbf{0}$ and $\chi_{ae} = \mathbf{0}$, this collapses to a more standard Latent Dirichlet Allocation.
- (ii). By having independent ‘attributes’, we reduce the state space.

³ Specifically, $J(1) = 2$ for the action; $J(2) = 4$ for source of authority, $J(3) = 5$ for justification and $J(4) = 3$ for tone.

⁴ That is, for the s th element of ϕ_{ka} – which we denote $\phi_{ka}^{(s)}$ – we have:

$$\phi_{ka}^{(s)} = \frac{\exp\left(\tilde{\phi}_{ka}^{(s)}\right)}{1 + \sum_{m=2}^{J(a)} \exp\left(\tilde{\phi}_{ka}^{(m)}\right)} \iff \tilde{\phi}_{ka}^{(s)} = \ln\left(\frac{\phi_{ka}^{(s)}}{1 - \sum_{m=2}^{J(a)} \phi_{ka}^{(m)}}\right).$$

⁵ The first element of $\tilde{\phi}_{ka}$, ψ_{av} , and χ_{ae} is normalised to zero. Moreover, we normalise the vignette and enumerator fixed effects by setting $\psi_{a1} = \mathbf{0}$ and $\chi_{a1} = \mathbf{0}$ for all a (reference levels).

⁶ ψ_{av} and χ_{ae} provide a measurement scheme; this is similar in spirit to Item Response Theory. That is, this captures the notion that different enumerators will have different underlying tendencies to assess different attributes, and similarly that for different vignettes respondents have different underlying tendencies to display traits. Note that, in the special case that $\psi_{av} = \mathbf{0}$ and $\chi_{ae} = \mathbf{0}$, this collapses to a more standard Latent Dirichlet Allocation.

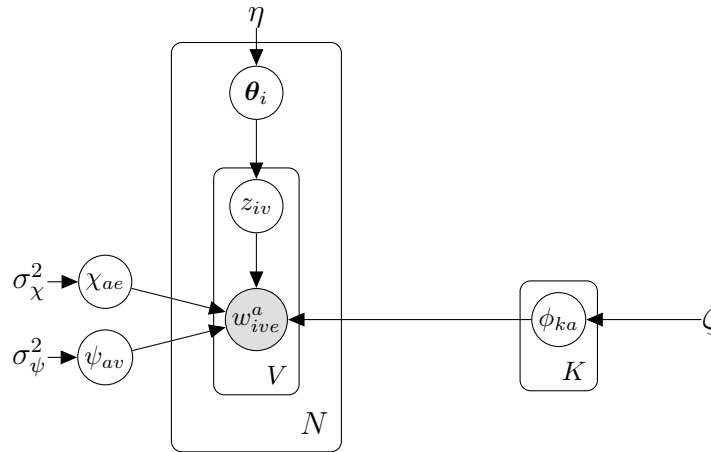
- (iii). Each attribute has a different “dictionary”.
- (iv). For each vector ψ_{av} and χ_{ae} the first element is normalised to zero.

We can then write the following generative model:

- (i). Draw ϕ_{ka} independently for $k \in \{1, \dots, K\}$ and for $a \in \{1, \dots, A\}$ from $\text{Dirichlet}(\zeta)$;
- (ii). Draw θ_i independently for $i \in \{1, \dots, N\}$ from $\text{Dirichlet}(\eta)$, where θ_i is a K -dimensional vector;
- (iii). Draw ψ_{av} and χ_{ae} from some suitable prior (respectively $\mathcal{N}(0, \sigma_\psi^2)$ and $\mathcal{N}(0, \sigma_\chi^2)$);
- (iv). For each vignette, draw a type z_{iv} from θ_i and draw an attribute from $w_{aive} \mid \phi_{z_{iv}, a}, \psi_{av}, \chi_{ae}$ (using the Multinomial Logit formula above).

This model can be described graphically as follows:

Figure A.1: Augmented Latent Dirichlet Allocation: Simplified plate diagram



Notes: Simplified plate diagram. For clarity we suppress plates over attributes $a \in \{1, \dots, A\}$ and enumerators $e \in \{1, 2\}$. Conditional on $(z_{iv}, \phi, \psi, \chi)$, the observed responses w_{ive}^a are independent across a and e .

The total probability of the model is as follows:⁷

$$p(W, \boldsymbol{\theta}, \boldsymbol{\phi} \mid \eta, \zeta, \boldsymbol{\psi}, \boldsymbol{\chi}) = \left[\prod_{k=1}^K \prod_{a=1}^A p(\boldsymbol{\phi}_{ka} \mid \zeta) \right] \left[\prod_{i=1}^N p(\boldsymbol{\theta}_i \mid \eta) \right] \\ \times \left[\prod_{i=1}^N \prod_{v=1}^V \sum_{k=1}^K p(z_{iv} = k \mid \boldsymbol{\theta}_i) \prod_{e=1}^2 \prod_{a=1}^A p(w_{ive}^a \mid \boldsymbol{\phi}_{ka}, \boldsymbol{\psi}_{av}, \boldsymbol{\chi}_{ae}) \right].$$

where $p(\boldsymbol{\phi}_{ka} \mid \zeta)$ and $p(\boldsymbol{\theta}_i \mid \eta)$ follow a latent dirichlet distribution. Conditional on $(\boldsymbol{\theta}_i, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\chi})$, the observed responses w_{ive}^a are generated from a K -component mixture: for each (i, v) , z_{iv} follows a categorical distribution and $w_{ive}^a \mid z_{iv} = k$ a multinomial logit distribution.

To fully specify the model, we need to impose an additional set of (hyper)priors; we choose:

$$\begin{aligned} \zeta &= 0.3; \\ \eta &= 1; \\ \boldsymbol{\chi}_{ae} &\sim \mathcal{N}(0, 3); \\ \boldsymbol{\psi}_{av} &\sim \mathcal{N}(0, 3). \end{aligned}$$

In practice, the model is highly insensitive to the choice of the (symmetric) hyperparameter ζ , as the data is highly informative for the parameter $\boldsymbol{\phi}_k$. However, the model is more sensitive to the choice of the (symmetric) hyperparameter η since the parameter $\boldsymbol{\theta}_i$ is estimated based only on data from a single candidate. The primary effect of the choice of η is the distance to the edge of the simplex of resulting parameter estimates.

E.2 Plackett-Luce model

To analyse the ranking data from the HR managers we develop the following Plackett-Luce model. The model employs a Bayesian hierarchical structure where preferences for types and attributes can vary across firms. We include an individual random effect for each individual candidate to improve the precision of the resulting estimates:

$$U_{fiv} = \boldsymbol{\beta}_f \boldsymbol{\theta}_i + \boldsymbol{\alpha}_f x_i + \gamma_i + \varepsilon_{fiv}, \quad (2)$$

⁷ For clarity of exposition we omit the normal prior for $\boldsymbol{\psi}$ and $\boldsymbol{\chi}$ from this equation. These enter the log-likelihood linearly.

where we have the following explanatory variables:

$$\begin{aligned} \theta_i &\implies && \text{a latent parameter from the Dirichlet model; } \tilde{\theta}_i \\ x_i &\implies && \text{“Known”}. \end{aligned}$$

We specify the error term as follows:

$$\varepsilon_{fiv} \sim EV1$$

Then we employ a Bayesian hierarchical structure for each set of parameters. For the individual mean γ_i , where we specify the error structure for the standard deviation on a bounded uniform distribution; we specify:⁸

$$\begin{aligned} \gamma_i &\sim \mathcal{N}(0, \sigma_\gamma^2); \\ \sigma_\gamma &\sim U(0, 10). \end{aligned}$$

For identification, we set $\gamma_i = 0$ for the first member of every triplet. Then for the α and β parameters we use the following structure:

$$\begin{aligned} \alpha_f &\sim \mathcal{N}(\alpha, \sigma_\alpha^2); \\ \alpha &\sim \mathcal{N}(0, 3); \\ \sigma_\alpha &\sim U(0, 10). \end{aligned}$$

and

$$\begin{aligned} \beta_f &\sim \mathcal{N}(\beta, \sigma_\beta^2); \\ \beta &\sim \mathcal{N}(0, 3); \\ \sigma_\beta &\sim U(0, 10). \end{aligned}$$

This implies the following total probability (denoting by θ the parameters, by y the data, and by λ the hyperparameters, and denoting $U(\beta_f, \alpha_f, \gamma_i, \theta_i, \mathbf{x}_i) \equiv U_{fiv}$) using the superscript r to denote the rank the

⁸ In setting the prior for the standard deviation as a uniform distribution we follow [Gelman \(2006\)](#).

HR manager assigns within the triplet (with 1 being the highest rank):

$$P(y|\boldsymbol{\beta}_f, \boldsymbol{\alpha}_f, \gamma_i, \boldsymbol{\theta}_i, \mathbf{x}_i) = \prod_{f=1}^F \prod_{v=1}^5 \frac{\exp(U_{fiv}^1)}{\sum_{r=1}^3 \exp(U_{fiv}^r)} \cdot \frac{\exp(U_{fiv}^2)}{\sum_{r=2}^3 \exp(U_{fiv}^r)},$$

We then specify the distribution of α_f , γ_i and β_f as:

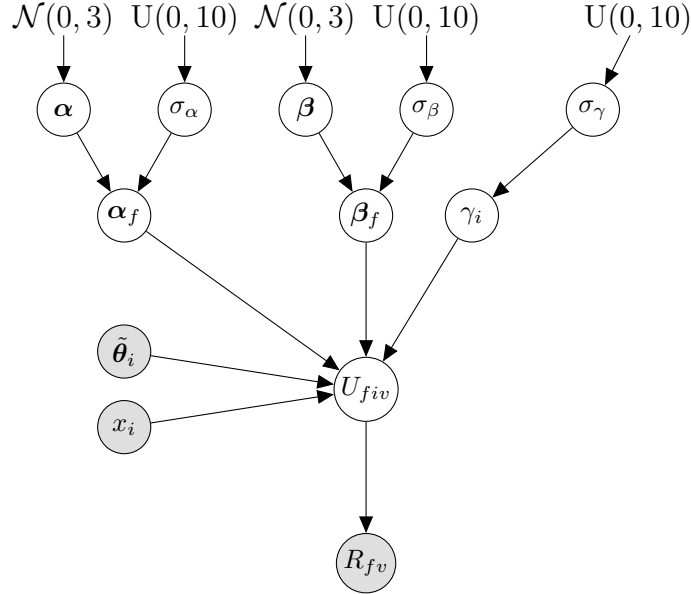
$$P(\boldsymbol{\beta}_f, \boldsymbol{\alpha}_f, \gamma_i | \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\gamma, \sigma_\alpha, \sigma_\beta) = \prod_{i=1}^N \mathcal{N}(\gamma_i | 0, \sigma_\gamma) \prod_{f=1}^F \mathcal{N}(\boldsymbol{\alpha}_f | \boldsymbol{\alpha}, \sigma_\alpha) \cdot \mathcal{N}(\boldsymbol{\beta}_f | \boldsymbol{\beta}, \sigma_\beta).$$

Finally, we define the hyperprior as:

$$P(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\gamma, \sigma_\alpha, \sigma_\beta) = \mathcal{N}(\boldsymbol{\alpha} | 0, 3) \cdot \mathcal{N}(\boldsymbol{\beta} | 0, 3) \cdot U(\sigma_\alpha, 0, 10) \cdot U(\sigma_\beta, 0, 10) \cdot U(\sigma_\gamma, 0, 10).$$

The full probability of drawing a set of parameters is the product of these three elements (following Bayes' rule).

Figure A.2: Hierarchical Plackett–Luce: Simplified plate diagram



Notes: Simplified plate diagram. For clarity we suppress the alternative-within-choice-set index $r \in \{1, 2, 3\}$ (triplet members) and the vignette index where not needed; utilities are constructed for each alternative and fed into the Plackett–Luce likelihood.

E.3 Estimation

We estimate this model using Hamiltonian Monte Carlo (HMC), implemented in Stan (Gelman, Lee, and Guo, 2015; Neal and Neal, 1996; Carpenter, Gelman, Hoffman, Lee, Goodrich, Betancourt, Brubaker, Guo, Li, and Riddell, 2017). HMC is an advanced algorithm for Markov Chain Monte Carlo (MCMC) simulations. Its key feature is the use of the gradient of the posterior distribution in determining step sizes. This allows HMC to consider how the probability of the parameters changes, taking more informed steps during the simulation. This method is particularly useful because it often converges to high-probability regions faster and more efficiently than traditional MCMC methods. This is especially true in complex models with many parameters, like the one we implement. This efficiency is due to its ability to avoid getting stuck in less probable regions, a common issue with simpler MCMC methods.

E.3.1 The Dirichlet model

We first estimate the Dirichlet model, and use these results for Section 3. A few things are worth noting: Firstly, the commonly used Gibbs Sampler (Steyvers, Smyth, Rosen-Zvi, and Griffiths, 2004) is not applicable due to the additional parameters in our model. However, the small number of types and attributes makes a direct estimation procedure feasible. Secondly, a K -component mixture distribution is invariant to permutations of component labels, resulting in $K!$ modes in the posterior distribution of the mixture parameters. This can be problematic when using HMC, as the leapfrog integrator may attempt to jump between different configurations of the component labels. In practice, the distinct modes of our posterior distribution prevent label switching in our draws. Consequently, we can re-label the types for consistent labeling across chains before performing standard convergence tests for MCMC simulations (Gelman and Rubin, 1992). To make inference easier, in particular for the joint model, we first run four shorter chains which we then relabel to each have the same type labels. We then use the estimates from these four shorter chains to start four new chains, which we use – following burn-in – in our analysis.

E.3.2 The Plackett-Luce model and joint-estimation

We estimate the Plackett-Luce model using Hamiltonian Monte Carlo (HMC) in Stan, implementing a non-centered reparameterization of the hierarchical parameters. This technique results in faster convergence and increases the precision of the model estimates. In such a nonlinear model, incorporating uncertainty in the first-stage parameters—the parameters of the Latent Dirichlet Allocation (LDA) model—presents a challenge. Properly propagating this uncertainty to the second-stage model provides a more realistic representation of the model’s confidence in its predictions and estimates. To address this, we follow Battaglia, Christensen, Hansen, and Sacher (2024), jointly estimating the LDA and Plackett-Luce models. This joint estimation approach leverages the curvature of the log-likelihood to guide the level of

Table A.6: Convergence statistics for the Dirichlet model with four types

Parameter	Average PSRF	Highest PSRF	Average effective sample size	Lowest effective sample size
Theta	0.999	1.009	2000	247
Phi Action	1.010	1.011	738	661
Phi Tone	0.999	1.001	1980	1359
Phi Justification	1.001	1.004	1462	965
Phi Authority	1.002	1.006	1116	791
Chi Action	0.999	0.999	2000	2000
Chi Tone	0.998	0.998	2000	2000
Chi Justification	0.999	1.000	2000	1748
Chi Authority	1.001	1.001	2000	1863
Psi Action	1.008	1.013	831	676
Psi Tone	0.999	1.002	2000	2000
Psi Justification	0.999	1.003	1777	1043
Psi Authority	1.001	1.006	1311	1228

Notes: Table A.6 shows the average and highest PSRF, and the average and lowest effective sample size for each subset of the parameters in the Dirichlet model. Crucially, the highest PSRF is low, showing good convergence of the model.

uncertainty about the type parameters, implicitly weighting the observations in the second-stage model by the uncertainty about θ_i .

Tables A.7 and A.8 show the potential scale reduction factor and average effective sample size of the chains. This is based on four chains with 1000 draws.

Table A.7: Convergence statistics for the joint model with four types for ranking as entry-level manager

Parameter	Average PSRF	Highest PSRF	Average effective sample size	Lowest effective sample size
Theta	1.000	1.007	4000	409
Phi Action	1.006	1.007	845	777
Phi Tone	1.000	1.002	3223	1891
Phi Justification	1.007	1.012	583	358
Phi Authority	1.001	1.004	1937	1298
Chi Action	0.999	0.999	4000	4000
Chi Tone	0.999	1	3496	3431
Chi Justification	0.999	1	4000	4000
Chi Authority	0.999	0.999	3268	3216
Psi Action	1.005	1.007	1013	870
Psi Tone	1.001	1.001	3220	2595
Psi Justification	1.000	1.011	2626	379
Psi Authority	1.000	1.003	1987	1776
Beta	1.004	1.004	1115	1066
Beta FS	0.999	1.006	4000	771
Beta Std	1.028	1.054	352	180
Gamma	0.999	1.001	4000	1349
Gamma Std	1.001	1.001	1232	1232

Notes: This table reports the average and highest PSRF, and the average and lowest effective sample size for each subset of the parameters in the jointly estimated model for the rankings as potential entry-level managers. Crucially, the highest PSRF is low, showing good convergence of the model.

Table A.8: Convergence statistics for the joint model with four types for ranking as entrepreneur

Parameter	Average PSRF	Highest PSRF	Average effective sample size	Lowest effective sample size
Theta	1.000	1.008	4000	387
Phi Action	1.006	1.008	1022	943
Phi Tone	1.001	1.004	2449	1475
Phi Justification	1.018	1.027	419	241
Phi Authority	1.002	1.007	1746	993
Chi Action	0.999	0.999	4000	4000
Chi Tone	1.002	1.003	2709	2662
Chi Justification	1.000	1.001	4000	4000
Chi Authority	1.000	1.000	2821	2814
Psi Action	1.005	1.009	1080	947
Psi Tone	1.002	1.004	2349	1960
Psi Justification	1.001	1.027	2619	243
Psi Authority	1.003	1.006	1528	1262
Beta	1.006	1.006	1124	1075
Beta FS	0.999	1.003	4000	672
Beta Std	1.023	1.027	252	182
Gamma	0.999	1.002	4000	846
Gamma Std	1.002	1.002	1285	1285

Notes: This table reports the average and highest PSRF, and the average and lowest effective sample size for each subset of the parameters in the jointly estimated model for the rankings as potential entrepreneurs. Crucially, the highest PSRF is low, showing good convergence of the model.

E.3.3 Stability of type parameters with joint estimation

Figures A.3 and A.4 show that the estimated type probability and distribution over type is consistent in the individually estimated Dirichlet model and the jointly estimated model.

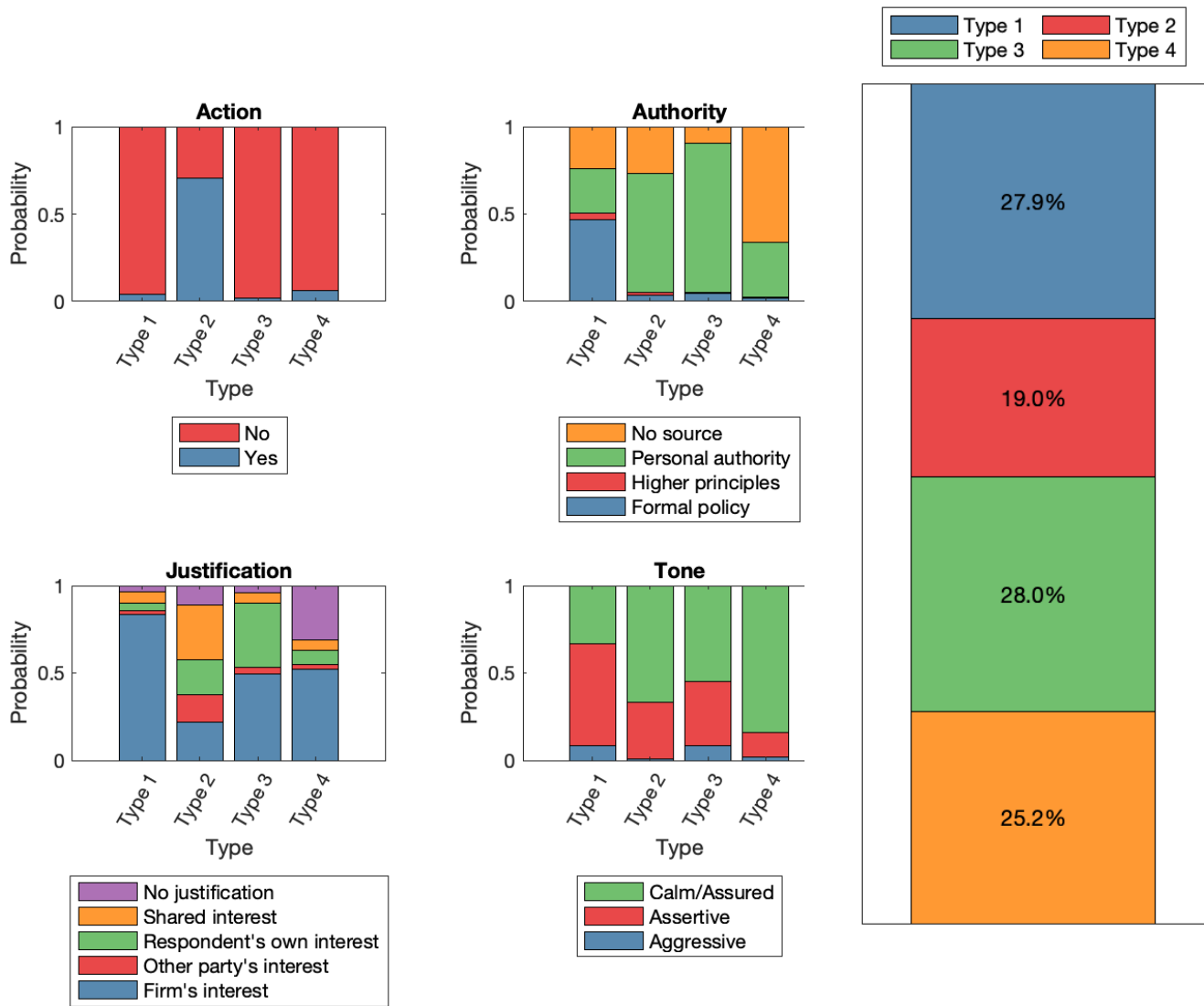


Figure A.3: The type parameters and distribution over types in the Dirichlet model

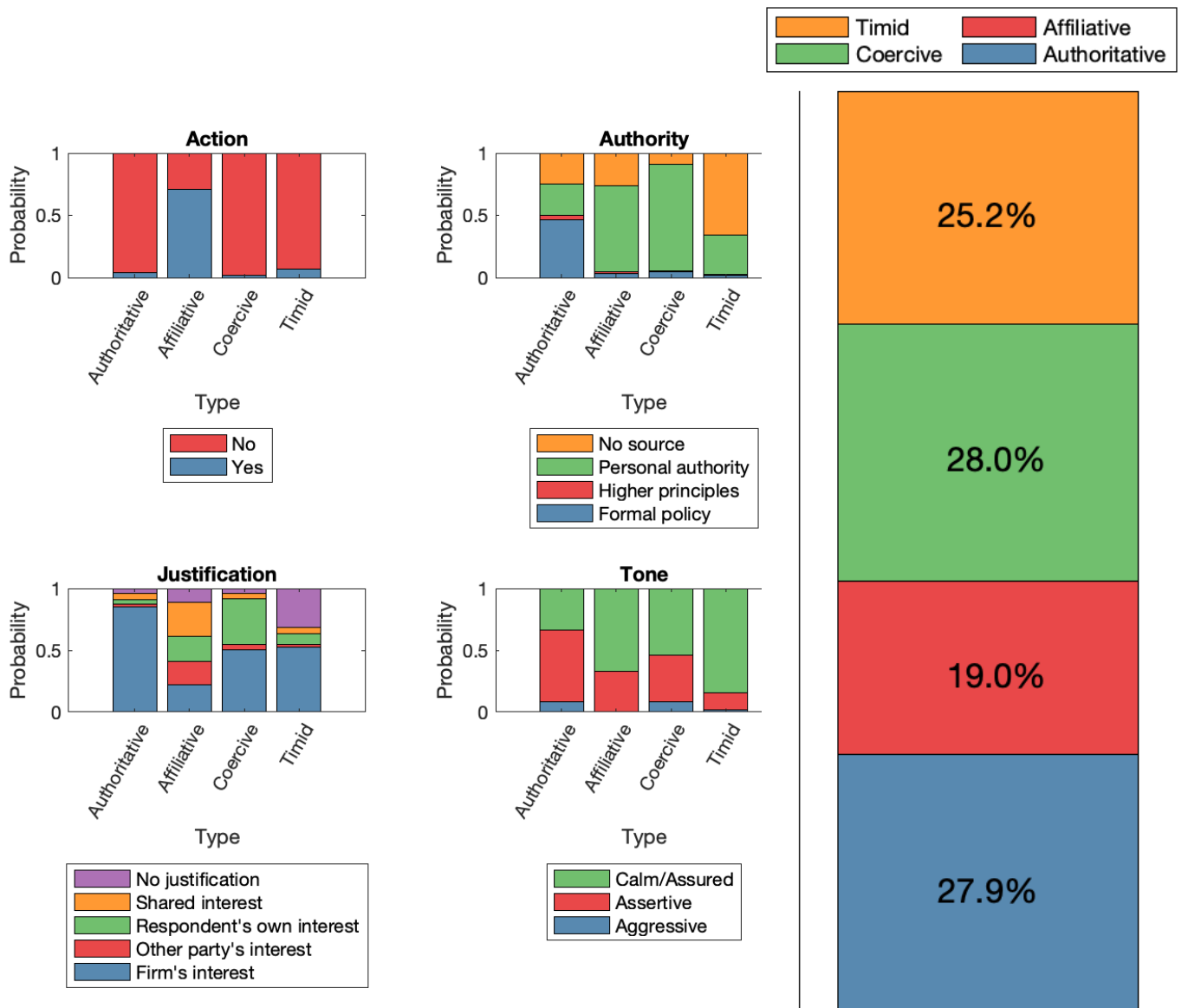


Figure A.4: The type parameters and distribution over types in the joint model

F Sensitivity to the choice of number of types

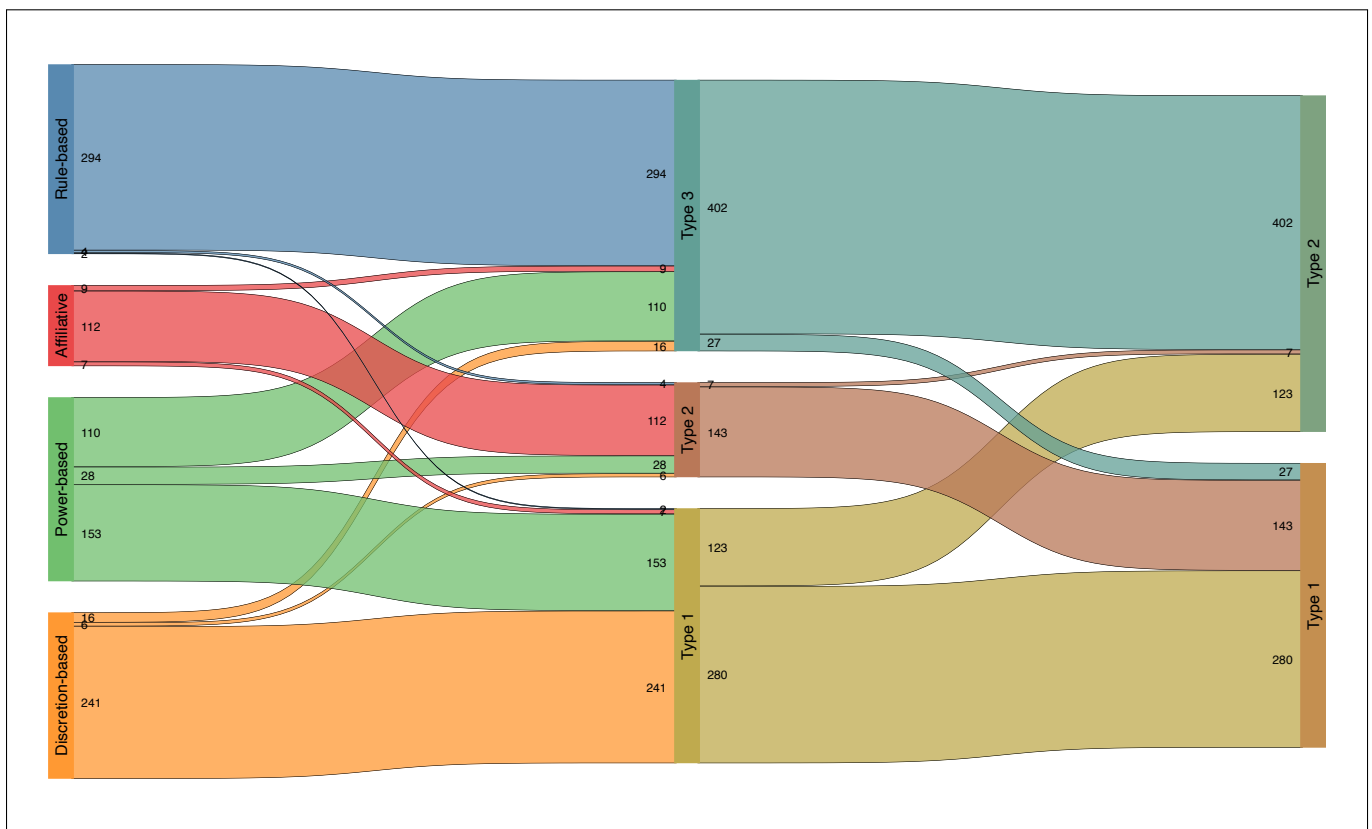
Our preferred estimation uses $K = 4$ types. In this appendix section, we consider alternative specifications with two, three and five types. In short, we argue that (i) our results are robust to using $K = 2$ or $K = 3$, and (ii) the model fit becomes unstable for $K = 5$ (in a manner that we explain shortly), and the model with $K = 5$ does not meaningfully improve on the fit compared to $K = 4$.

F.1 Considering two types and three types

On the following pages, we repeat each of the key estimations from our preferred model, but using $K = 2$ and then $K = 3$. By way of summary, the following key results of the paper are replicated across these two alternative specification:

- (i). In terms of attributes of the responses, we find an ‘affiliative’-like and ‘rule-based’-like type in all specifications. For $K = 2$, these are essentially the two types that we find. For $K = 3$, we have an affiliative type, a rule-based type and a power/discretion-based type. When $K = 4$, the latter two types are split into an rule-based, power-based and discretion-based type. We illustrate this in Figure A.5. This figure shows the key intuition for why our results are stable across the choice of types: each specification preserves the essential distinction between the rule-based type and the other type(s).
- (ii). For $K = 2$, $K = 3$ and $K = 4$, the rule-based type is in each case more likely to be male, has better labour market outcomes, and is better able to predict the decision of the senior manager.
- (iii). For $K = 2$, $K = 3$ and $K = 4$, the management experience treatment causes candidates with low parental education to be more likely to be rule-based. For $K = 3$ and $K = 4$, we learn that this is driven by a shift from a power-based to an rule-based management style (an insight that is not possible for the $K = 2$ model).
- (iv). For $K = 2$, $K = 3$ and $K = 4$, the rule-based type is strongly preferred by firm managers.
- (v). The effect of the first actors’ gender on the realised type is comparable across specifications, but most pronounced when $K = 4$.

Figure A.5: Sankey flow diagram: Assignment of types under $K = 2$, $K = 3$ and $K = 4$



F.1.1 Results: Two-type model

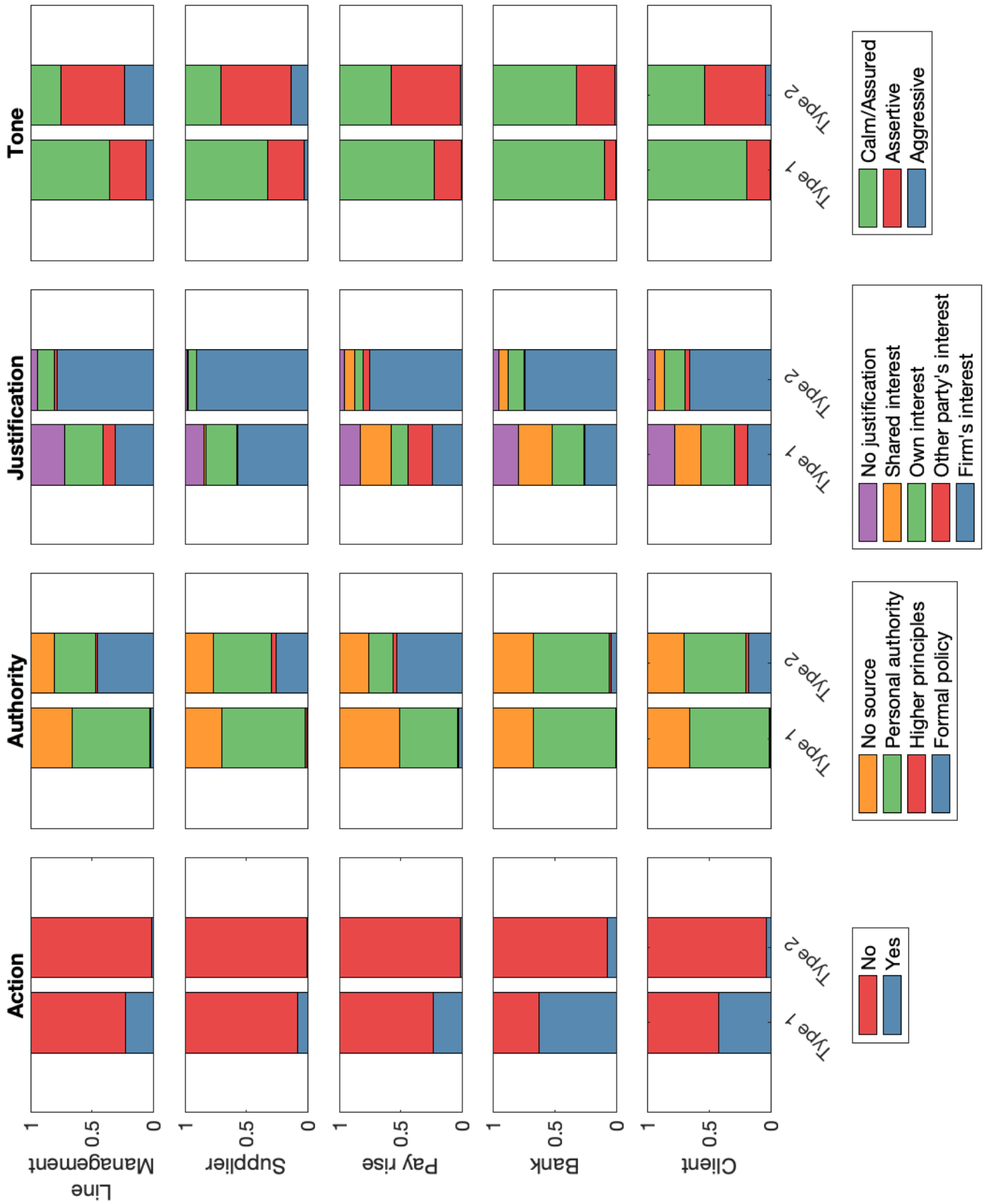


Figure A.6: 'pure type' management styles amongst Ethiopian young professionals

Figure A.7: Distribution of types across individuals

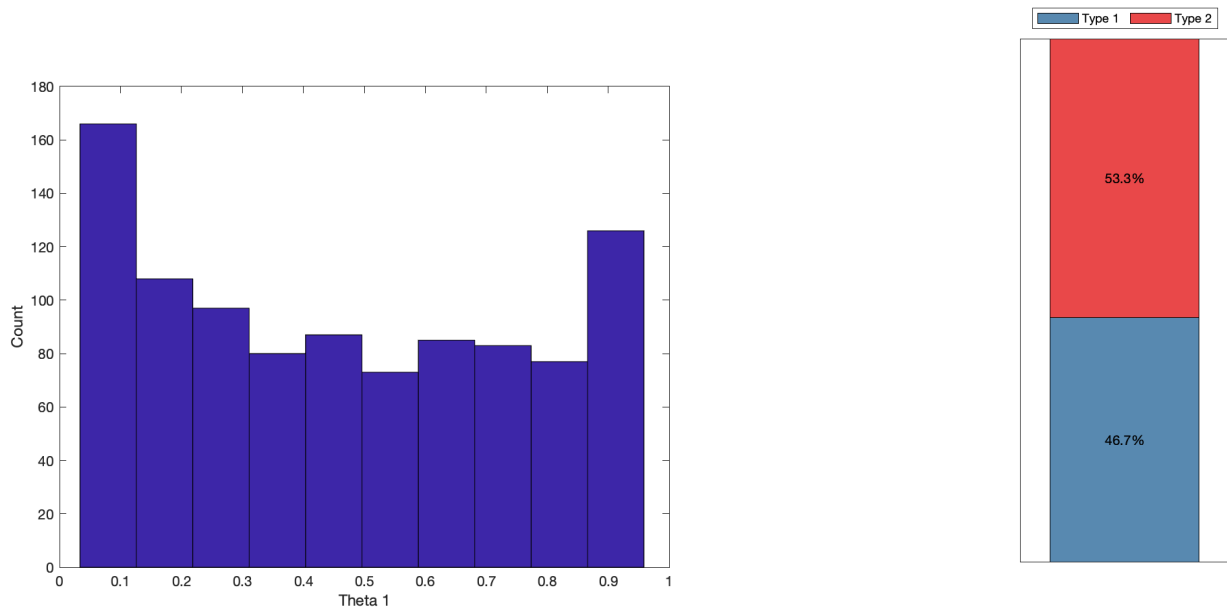


Table A.9: Agreement in rankings between the respondents' and the HR consultant's rankings.

	Prediction Senior manager and senior manager	Own action and Senior Manager	Internal Agreement
Overall	.511	.415	.793
Type 1	.464	.382	.814
Type 2	.551	.443	.775

Notes This table displays Kendall's Tau values comparing the respondents' rankings and the expert's rankings. The column "Prediction of Senior Manager versus Actual Senior Manager" compares the respondents' predictions of the senior manager's actions with the actual senior manager's actions. The column "Own Action versus the Senior Manager" compares the senior manager's actions with what the respondent would do themselves. The column "Internal Agreement" displays Kendall's Tau for the respondents' perception of their own actions versus what they expect a senior manager would do. Kendall's Tau is calculated as $\tau = \frac{\sum \text{agreements} - \sum \text{disagreements}}{\sum \text{pairwise comparisons}}$.

Table A.10: Characteristics and Types: Summary Statistics

Type	Gender [1=male]	Wage employment indicator	Self- Employment indicator	Above Median Reservation Wage	Above Median Reservation Profit	Average Duration Response
Type 1	.781	.668	.149	.287	.300	36.0
Type 2	.818	.699	.174	.404	.408	50.5
<i>p</i> -value	.038	.086	.29	<0.001	<0.001	<0.001

Notes This table describes the average characteristics of individuals of each type. Specific, this includes their gender (1 indicating male, 0 female), a dummy for their wage- and self-employment status, and the probability they have an above-median reservation wage and profit based on data collected before respondents' attended the studio. We also include a dummy for whether or not the individual was treated in the management placement experiment, and the average duration of the responses of the candidate across the vignettes. Note that the median splits on reservation wage and profit do not yield a 50/50 split due to bunching in the underlying data at ETB10.000. These numbers are calculated by assigning each individual to the pure type for which they have the highest estimated $\hat{\theta}_i$. Then, a conditional average is taken for each pure type. To test whether intentions differ across types, we compute the Mahalanobis (Wald) statistic and compare it to a χ^2_3 distribution. This relies on the multivariate normal approximation to the posterior described in Section 4.1 of [Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin \(2013\)](#). The resulting *p*-value appears in the final row. For non-binary outcomes, we first create a binary split at the median to split the sample.

Table A.11: Labour market experience and Types: Summary Statistics

Type	Employment <i>Years</i>	Permanent employment <i>Years</i>	Unemployed <i>Years</i>	Management Position <i>Share</i>	Number of transitions <i>Count</i>
Type 1	5.217	3.564	.525	.117	.942
Type 2	5.417	3.801	.417	.168	.735
<i>p</i> -value	.058	.009	.022	.002	.006

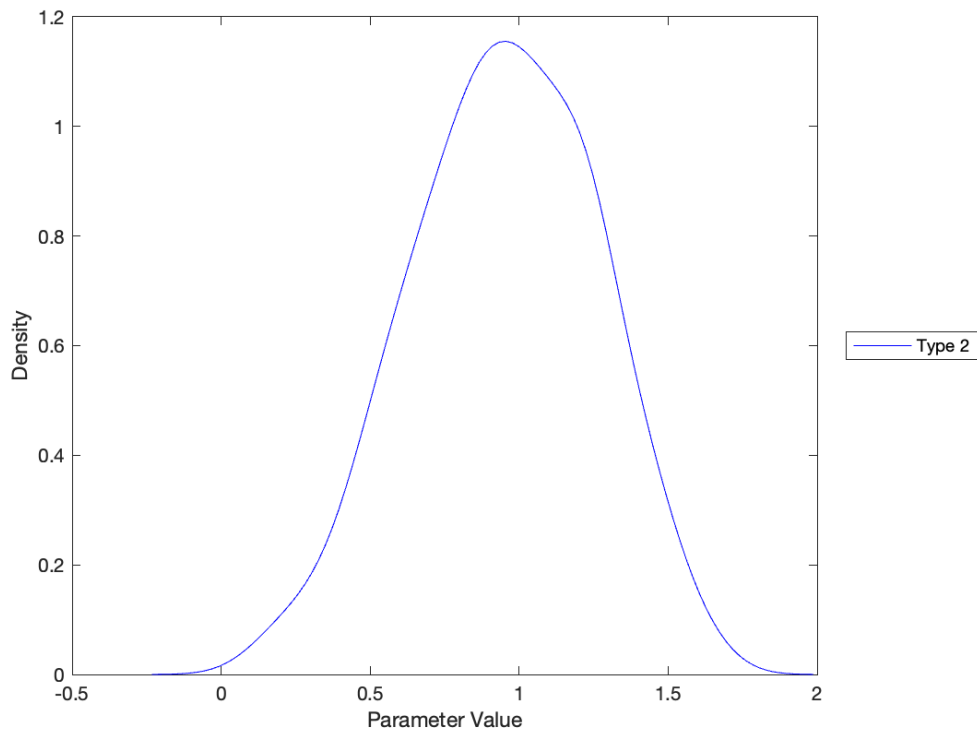
Notes This table describes the labour market experience over the past 6 years of individuals of each type. Specific, this includes the number of years they have been in employment including both self- and wage-employment, the number of years they have been in permanent employment, the number of years they have been unemployed, whether they were in a management position before participating in the studio experiment, and finally the number of labour market transitions they have gone through. These numbers are calculated by assigning each individual to the pure type for which they have the highest estimated $\hat{\theta}_i$. Then, a conditional average is taken for each pure type. To test whether intentions differ across types, we compute the Mahalanobis (Wald) statistic and compare it to a χ^2_3 distribution. This relies on the multivariate normal approximation to the posterior described in Section 4.1 of Gelman et al. (2013). The resulting *p*-value appears in the final row. For non-binary outcomes, we first create a binary split at the median to split the sample.

Table A.12: The causal effect of managerial experience on management style by parents education

	Full sample		Low parental education		High parental education	
	(1)	(2)	(3)	(4)	(5)	(6)
Type 1 (%)	47.9	-2.2	52.3	-4.5	43.9	0.0
		[-4.7, 1.0]		[-8.3, -0.7]		[-3.5, 3.4]
Type 2 (%)	52.1	2.2	47.7	4.5	56.1	-0.0
		[-1.0, 4.7]		[0.7, 8.3]		[-3.4, 3.5]
N	479	500	229	239	250	261

Notes This table reports the treatment effect of the management experience experiment on the managerial traits of individuals. The treatment effect is calculated based on the distribution of the difference in the average value of θ for treated and untreated individuals. Columns (1), (3) and (5) report the average estimated value of θ_i for individuals that were not treated in the management experience experiment for respectively all individuals, individuals whose parents did not finish primary school and for individuals for whom at least one parent did. Columns (2), (4) and (6) report the treatment effect of the management experience experiment on their managerial traits for these three groups respectively. In columns (2), (4) and (6) both the average treatment effect and the 95% credible interval, in square brackets, are reported.

Figure A.8: The distribution of the preferences for entry-level managers



Notes This figure shows the distribution of the estimates for $\hat{\beta}_{fs}$ across all firms in terms of demand for entry-level managers. In this figure, the Type 1 is omitted and the estimates for $\hat{\beta}_{\text{Type 2}}$, $\hat{\beta}_{\text{Type 3}}$ are plotted

Table A.13: Actors' gender and management styles

Panel A: Effect on estimated types				
	Theta 1	Theta 2		
First actor female	-0.022	0.022		
Constant	0.478***	0.522***		
Bayesian Credible Interval	[-.046 .004]	[-.004 .046]		
N	982	982		
Panel B: Effect on attributes				
	Justification		Authority	
	Rely on firm's interest	Rely on own interest	Rely on formal authority	Rely on seniority
First actor female	0.032* (0.019)	-0.028** (0.014)	0.021 (0.015)	-0.026 (0.020)
Constant	0.554*** (0.013)	0.178*** (0.010)	0.154*** (0.010)	0.537*** (0.014)
Enumerator FE	Yes	Yes	Yes	Yes
Vignette FE	Yes	Yes	Yes	Yes
Mean dep. var	0.552	0.168	0.163	0.511
N	6887	6887	6887	6887
Panel C: Effect on managers' assessments				
	Ranking Data		Normalised likert score	
	Manager	Entrepreneur	Manager	Entrepreneur
First actor female	0.052** [0.010 0.093]	0.060*** [0.017 0.102]	0.083** (0.033)	0.093*** (0.033)
Constant	0.474*** [0.453 0.493]	0.470*** [0.449 0.492]	-0.045* (0.025)	-0.050** (0.025)
Vignette FE			Yes	Yes
N			6874	6869

Notes: This figures displays the causal link between the first actor a respondent sees and their subsequent responses (for the second to fifth vignette). Appendix L reports a sequence of multinomial logit regressions on the full set of attributes from which those in panel B were selected. The third panel shows the effect of (a) the first actor being female and (2-3) the results from an Acharya-style mediation analysis for using first the estimates for the type parameter from Panel B) and then the attributes of the responses from Panel A. Statistical significance is denoted by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

F.1.2 Results: Three-type model

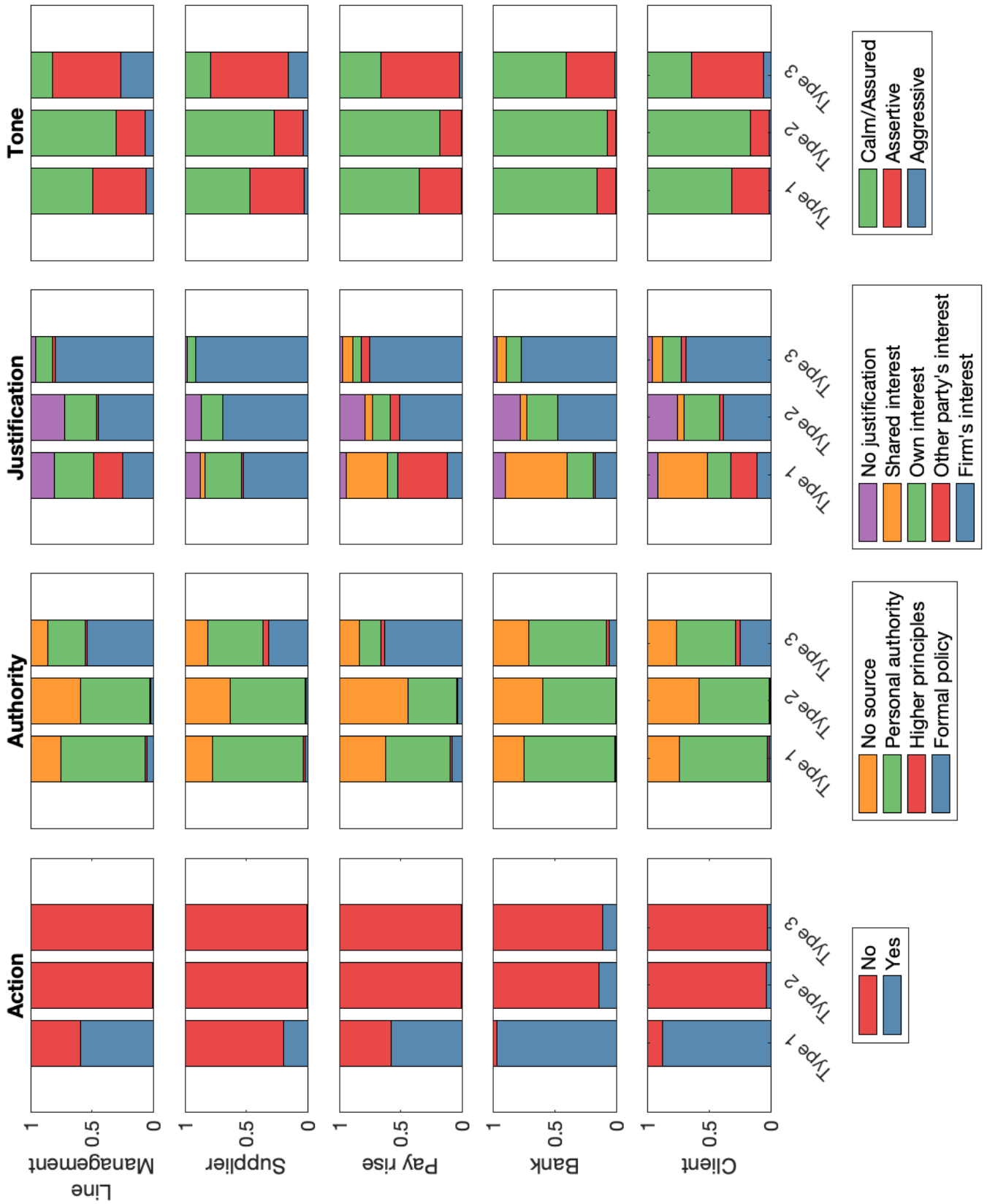


Figure A.9: 'pure type' management styles amongst Ethiopian young professionals

Figure A.10: Distribution of types across individuals

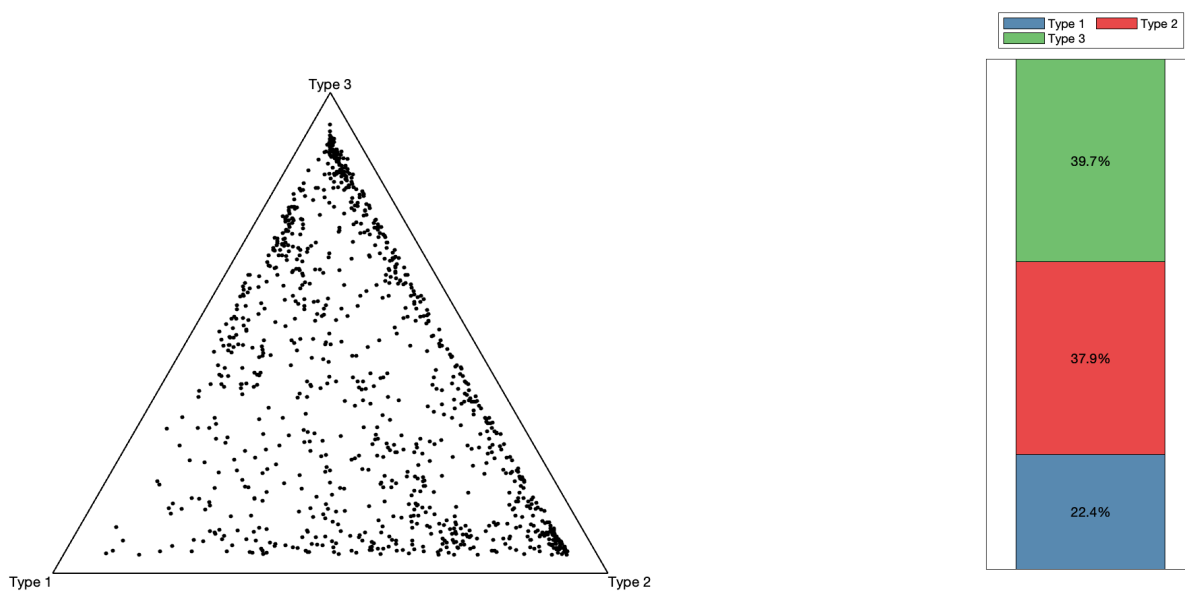


Table A.14: Agreement in rankings between the respondents' and the HR consultant's rankings.

	Prediction Senior manager and senior manager	Own action and Senior Manager	Internal Agreement
Overall	.511	.415	.793
Type 1	.502	.414	.811
Type 2	.457	.376	.812
Type 3	.565	.451	.768

Notes This table displays Kendall's Tau values comparing the respondents' rankings and the expert's rankings. The column "Prediction of Senior Manager versus Actual Senior Manager" compares the respondents' predictions of the senior manager's actions with the actual senior manager's actions. The column "Own Action versus the Senior Manager" compares the senior manager's actions with what the respondent would do themselves. The column "Internal Agreement" displays Kendall's Tau for the respondents' perception of their own actions versus what they expect a senior manager would do. Kendall's Tau is calculated as $\tau = \frac{\sum \text{agreements} - \sum \text{disagreements}}{\sum \text{pairwise comparisons}}$.

Table A.15: Characteristics and Types: Summary Statistics

Type	Gender [1=male]	Wage employment indicator	Self- Employment indicator	Above Median Reservation Wage	Above Median Reservation Profit	Average Duration Response
Type 1	.772	.676	.154	.34	.31	38.4
Type 2	.784	.671	.143	.266	.312	36.7
Type 3	.83	.701	.185	.436	.425	53.1
<i>p</i> -value	.042	.476	.152	<0.001	<0.001	< 0.001

Notes This table describes the average characteristics of individuals of each type. Specific, this includes their gender (1 indicating male, 0 female), a dummy for their wage- and self-employment status, and the probability they have an above-median reservation wage and profit based on data collected before respondents' attended the studio. We also include a dummy for whether or not the individual was treated in the management placement experiment, and the average duration of the responses of the candidate across the vignettes. Note that the median splits on reservation wage and profit do not yield a 50/50 split due to bunching in the underlying data at ETB10.000. These numbers are calculated by assigning each individual to the pure type for which they have the highest estimated $\hat{\theta}_i$. Then, a conditional average is taken for each pure type. To test whether intentions differ across types, we compute the Mahalanobis (Wald) statistic and compare it to a χ^2_3 distribution. This relies on the multivariate normal approximation to the posterior described in Section 4.1 of Gelman et al. (2013). The resulting *p*-value appears in the final row. For non-binary outcomes, we first create a binary split at the median to split the sample.

Table A.16: Labour market experience and Types: Summary Statistics

Type	Employment <i>Years</i>	Permanent employment <i>Years</i>	Unemployed <i>Years</i>	Management Position <i>Share</i>	Number of transitions <i>Count</i>
Type 1	5.292	3.629	.466	.121	.879
Type 2	5.178	3.545	.573	.114	.955
Type 3	5.484	3.864	.363	.185	.688
<i>p</i> -value	.053	.015	.002	.002	.021

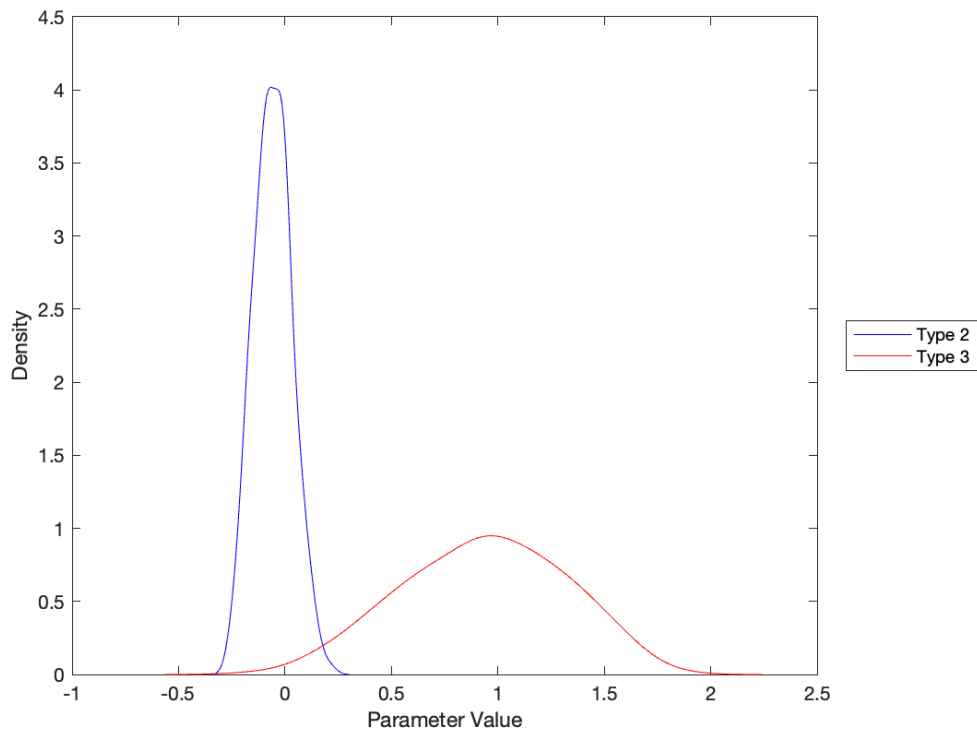
Notes This table describes the labour market experience over the past 6 years of individuals of each type. Specific, this includes the number of years they have been in employment including both self- and wage-employment, the number of years they have been in permanent employment, the number of years they have been unemployed, whether they were in a management position before participating in the studio experiment, and finally the number of labour market transitions they have gone through. These numbers are calculated by assigning each individual to the pure type for which they have the highest estimated $\hat{\theta}_i$. Then, a conditional average is taken for each pure type. To test whether intentions differ across types, we compute the Mahalanobis (Wald) statistic and compare it to a χ^2_3 distribution. This relies on the multivariate normal approximation to the posterior described in Section 4.1 of Gelman et al. (2013). The resulting *p*-value appears in the final row. For non-binary outcomes, we first create a binary split at the median to split the sample.

Table A.17: The causal effect of managerial experience on management style by parents education

	Full sample		Low parental education		High parental education	
	(1)	(2)	(3)	(4)	(5)	(6)
Type 1 (%)	21.9	1.4	22.9	1.1	21.1	1.6
		[-0.7, 3.6]		[-2.0, 4.0]		[-1.2, 4.4]
Type 2 (%)	40.4	-4.6	43.7	-7.0	37.3	-2.4
		[-6.8, -1.9]		[-10.7, -3.4]		[-5.8, 0.8]
Type 3 (%)	37.7	3.2	33.4	5.8	41.6	0.8
		[0.6, 5.9]		[2.1, 9.4]		[-2.4, 4.0]
N	479	500	229	239	250	261

Notes This table reports the treatment effect of the management experience experiment on the managerial traits of individuals. The treatment effect is calculated based on the distribution of the difference in the average value of θ for treated and untreated individuals. Columns (1), (3) and (5) report the average estimated value of θ_i for individuals that were not treated in the management experience experiment for respectively all individuals, individuals whose parents did not finish primary school and for individuals for whom at least one parent did. Columns (2), (4) and (6) report the treatment effect of the management experience experiment on their managerial traits for these three groups respectively. In columns (2), (4) and (6) both the average treatment effect and the 95% credible interval, in square brackets, are reported.

Figure A.11: The distribution of the preferences for entry-level managers



Notes This figure shows the distribution of the estimates for $\hat{\beta}_{f_s}$ across all firms in terms of demand for entry-level managers. In this figure, the Type 1 is omitted and the estimates for $\beta_{\text{Type 2}}$, $\beta_{\text{Type 3}}$ are plotted

Table A.18: Actors' gender and management styles

Panel A: Effect on estimated types				
	Theta 1	Theta 2	Theta 3	
First actor female	-0.012	-0.009	0.020	
Constant	0.232***	0.385***	0.384***	
Bayesian Credible Interval	[-.028 .017]	[-.042 .007]	[0.000 .046]	
N	982	982	982	
Panel B: Effect on attributes				
	Justification		Authority	
	Rely on firm's interest	Rely on own interest	Rely on formal authority	Rely on seniority
First actor female	0.032*	-0.028**	0.021	-0.026
	(0.019)	(0.014)	(0.015)	(0.020)
Constant	0.554***	0.178***	0.154***	0.537***
	(0.013)	(0.010)	(0.010)	(0.014)
Enumerator FE	Yes	Yes	Yes	Yes
Vignette FE	Yes	Yes	Yes	Yes
Mean dep. var	0.552	0.168	0.163	0.511
N	6887	6887	6887	6887
Panel C: Effect on managers' assessments				
	Ranking Data		Normalised likert score	
	Manager	Entrepreneur	Manager	Entrepreneur
First actor female	0.052**	0.060***	0.083**	0.093***
	[0.010 0.093]	[0.017 0.102]	(0.033)	(0.033)
Constant	0.474***	0.470***	-0.045*	-0.050**
	[0.453 0.493]	[0.449 0.492]	(0.025)	(0.025)
Vignette FE			Yes	Yes
N			6874	6869

Notes: This figures displays the causal link between the first actor a respondent sees and their subsequent responses (for the second to fifth vignette). Appendix L reports a sequence of multinomial logit regressions on the full set of attributes from which those in panel B were selected. The third panel shows the effect of (a) the first actor being female and (2-3) the results from an Acharya-style mediation analysis for using first the estimates for the type parameter from Panel A) and then the attributes of the responses from Panel A. Statistical significance is denoted by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

F.2 Considering five types

Finally, we consider setting $K = 5$. There are several indications that five types are excessive in this context. First, the five-type model is bimodal – in the sense that we can find two distinct kinds of classification that have approximately equal log-likelihoods (*i.e.* bimodality in the substantive description of types – not merely bimodality due to ‘label-switching’). Second, one of these two alternative classifications is descriptively very similar to the results from the four-type model: in essence, the five types could be described as ‘rule-based’, ‘affiliative’, ‘power-based’, ‘discretion-based’ and ‘discretion-power-based’ – and, of these, the ‘discretion-based’ and ‘discretion-power-based’ behave in extremely similar ways. Third, neither of these alternative classifications seems to add any additional nuance or insight to the conclusions reached from the four-type model.

G Inclusion in the studio

Table A.19: Logit Regression Results on Studio Attendance and Management Experience Treatment

	(1) Baseline sample	(2) 4-year follow-up sample	(3) Studio sample
	Attended Studio	Attended Studio	Treated
Treated	0.0208 (0.0239)	0.0159 (0.0241)	
Gender (1=Female)	-0.138*** (0.0275)	-0.121*** (0.0283)	0.0783* (0.0413)
Age at baseline	-0.00425 (0.00435)	-0.000840 (0.00456)	-0.00149 (0.00589)
BA Degree	-0.109* (0.0591)	-0.0175 (0.0545)	0.0142 (0.0706)
Either parent finished primary school	-0.0135 (0.0243)	-0.000208 (0.0252)	-0.0128 (0.0327)
Either parent owned a business	-0.0311 (0.0247)	-0.0123 (0.0252)	-0.00308 (0.0333)
In wage employment		0.00745 (0.0318)	0.0664 (0.0420)
Manages others while in wage emp.		0.0635 (0.0420)	-0.0261 (0.0508)
In self-employment		-0.0127 (0.0395)	0.0609 (0.0516)
Above median reservation wage		-0.0726** (0.0294)	0.00834 (0.0390)
Above median reservation profit		0.0470 (0.0299)	0.0346 (0.0383)
Observations	1637	1429	993

Notes: This table reports the average marginal effects from a logit regression on attending the studio conditional of being part of the previous management experience experiment in column (1) and conditional on being part of the four-year follow-up in column (2). Finally, column (3) reports the average marginal effects from a logit regression on being treated conditional on attending the studio in column (2). As independent variable a set of variables related to socioeconomic background of studio participants, and labour market outcomes before attending the studio are included. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.20: Inverse Probability Weighted Mean of θ

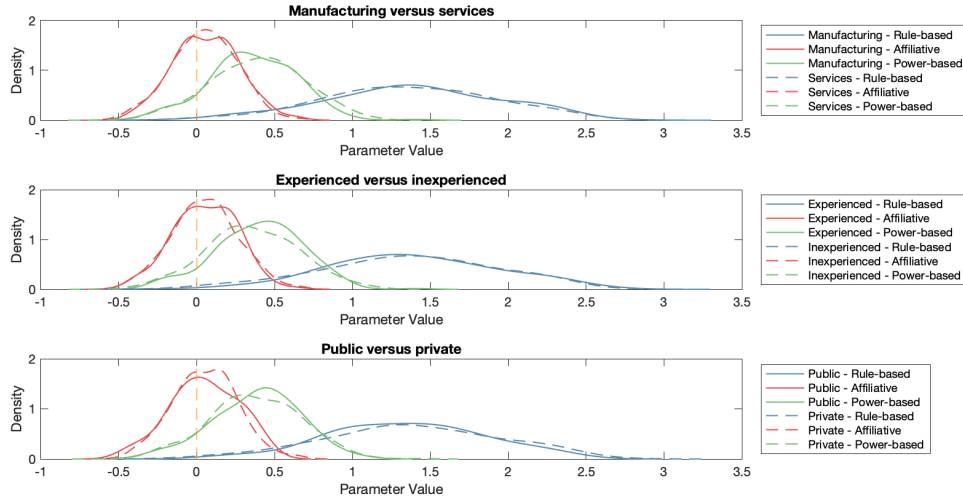
	Unweighted	Weighted
Discretion-based	.252	.252
Rule-based	.278	.276
Affiliative	.190	.190
Power-based	.280	.281

This table reports the unweighted mean of θ , and the mean of θ after an inverse probability weighting. This result suggests that the individuals included in the studio sample are representative of the full sample of individuals that were invited to participate in the studio experiment. The weights are calculated based on the probability of attending the studio as a function of a set of baseline characteristics from before the management experience experiment, including wage and self-employment status, gender, age, a dummy whether the individual has a bachelor's degree, a dummy for whether the individual has any intentions of starting a business, parents' education (a dummy equal to one if either parent finished primary school), and parents' experience running a business (a dummy equal to one if either parent ran a business).

H Variations on the Plackett-Luce model

H.1 The distribution of β_f conditional on covariates

Figure A.12: The distribution of β_f conditional on covariates



Notes This figure shows the distribution of the estimates for $\hat{\beta}_{fs}$ for a number of subsamples of the data. The top pane splits the sample by whether the firm operates in a manufacturing or services sector, the second pane splits the sample by the median number of years of experience of the manager, and the third pane splits the sample by whether the firm is public or private.

H.2 Allowing firm preferences to vary by vignette

$$U_{fiv} = \beta_{fv} \theta_i + \gamma_i + \varepsilon_{fiv}, \quad (3)$$

The priors for the Dirichlet model and γ_i are the same as in our main specification. For the parameters β_{fv} we specify the following priors:

$$\begin{aligned} \beta_{fv} &\sim \text{MVN}(\beta_v, \Sigma_v), \quad \text{with } \Sigma_v = \text{diag}(\sigma_v) \mathbf{I} \text{diag}(\sigma_v), \\ \beta_v &\sim \mathcal{N}(\mathbf{0}, (K-1) \mathbf{I}), \\ \sigma_{v,j} &\sim \text{Uniform}(0, 10) \quad \text{for } j = 1, \dots, K-1. \end{aligned}$$

Table A.21: Heterogeneity demand for types by vignette

	Employee Absence	Supplier	Pay Rise	Bank	Client
Rule-based	2.374 [1.552, 3.381]	1.803 [1.094, 2.671]	1.892 [1.235, 2.666]	1.304 [0.690, 1.994]	1.656 [0.976, 2.478]
Affiliative	0.254 [-0.533, 1.037]	0.035 [-0.809, 0.893]	-0.345 [-1.171, 0.472]	-0.038 [-0.810, 0.762]	0.008 [-0.767, 0.809]
Power-based	0.767 [0.076, 1.504]	-0.141 [-0.795, 0.497]	0.715 [0.020, 1.492]	0.059 [-0.498, 0.657]	0.737 [0.100, 1.450]

H.3 Allowing firm preferences to vary by candidate gender

$$U_{fiv} = \alpha_f \mathbb{I}[\text{female}]_i + \beta_f \theta_i + \delta \cdot \mathbb{I}[\text{female}]_i \cdot \theta_i + \gamma_i + \varepsilon_{fiv}, \quad (4)$$

The priors for the parameters α_f , β_f and γ_i are the same as in our main specification, as well as all priors in the Dirichlet model. For the new parameters, i.e. δ , we specify the following priors:

$$\delta \sim \mathcal{N}(0, 3);$$

Table A.22: Heterogeneity by the the gender of the candidate

Parameter	Estimate	90% Credible Interval
β_1	1.253	[0.925, 1.595]
β_2	0.205	[-0.254, 0.653]
β_3	0.286	[-0.039, 0.628]
α_1	-0.092	[-0.599, 0.410]
δ_1	0.692	[-0.055, 1.449]
δ_2	-0.457	[-1.396, 0.495]
δ_3	0.411	[-0.293, 1.141]

H.4 Overcontrolling for the duration of the response

We also re-estimate our main model controlling for the duration of the vignette to assess whether candidates with the rule-based managerial styles are preferred simply because they speak for longer. Of course, verbosity may itself reflect the type of behaviour we aim to capture, so this constitutes an over-control. Nonetheless, it serves as a useful robustness check for whether duration mechanically drives the results.

Specifically, we estimate the following equation, allowing the coefficient on duration, α , to vary by vignette. We place an independent normal prior on each element of α , with mean zero and variance three.

$$U_{fiv} = \alpha \cdot \text{duration}_{iv} + \beta_f \theta_i + \gamma_i + \varepsilon_{fiv}$$

Table A.23: Demand (over)controlling for duration

Parameter	Estimate	90% Credible Interval
$\beta_{rule-based}$	0.875	[0.567, 1.199]
$\beta_{affiliative}$	-0.012	[-0.394, 0.355]
$\beta_{power-based}$	0.231	[-0.059, 0.535]
α_1	0.010	[0.006, 0.013]
α_2	0.006	[0.002, 0.009]
α_3	0.009	[0.006, 0.012]
α_4	0.007	[0.003, 0.010]
α_5	0.008	[0.004, 0.012]

H.5 Causal effect of managerial experience

Table A.24: The causal effect of managerial experience on managers' assessments in log-odds space

Parameter	Estimate	90% CrI	
		Lower	Upper
β_1	0.054	-0.028	0.130
β_2	-0.190	-0.270	-0.110
β_3	0.150	0.035	0.260

Notes: This table reports the coefficients in mlogit space of a Plackett-Luce model with as Latent utility function $U_{fiv} = \beta_1 \text{treated} + \beta_2 \text{low parental education} + \beta_3 \text{treated} \cdot \text{low parental education} + \varepsilon_{fiv}$ where $\varepsilon_{fiv} \sim \text{Gumbel}(0, 1)$. Effective sample sizes N_{eff} were 1461 (β_1), 1530 (β_2), and 1418 (β_3). PSRF (\hat{R}) was 1.00 for all parameters.

I Describing heterogeneity in firm preferences

To assess how heterogeneous firms are in their revealed preferences for aspiring manager, we first assess the probability that two managers agree with each other, both for the full sample and by individual vignette. Table A.25 shows that the probability of agreement between two managers varies between 56% and 61%

across the vignettes. To quantify the level of agreement between firms, we implement a random-effects rank-ordered logit model without covariates; we specify the latent utility from firm f assessing candidate i as:

$$y_{fivs}^* = \mu_{ivs} + \varepsilon_{fivs},$$

where we specify $\mu_{ivs} \sim \mathcal{N}(0, \sigma_\mu^2)$ and we assume that ε_{fivs} has a Type 1 Extreme Value distribution.⁹ We assume that μ_{ivs} is independent across vignettes and types of employment (so, in effect, we estimate this model separately for each vignette and each type of employment).¹⁰ This model reveals much about firm preferences and the notion of ‘management as technology’ because the variance σ_μ^2 captures the strength of relative agreement among firms. In one limiting case, firm preferences are completely idiosyncratic: $\sigma_\mu^2 = 0$. In another limiting case, preferences are common across all firms: $\sigma_\mu^2 \rightarrow \infty$. An important intermediate case is where the proportion of variation due to common preferences equals the proportion due to idiosyncratic preferences: $\sigma_\mu^2 = \pi^2/6$. In this way, estimation of σ_μ^2 provides a specific figure — to our understanding, the first formal quantification in the literature — for heterogeneity in firm preferences over management traits.¹¹

We find that, in line with the probability of agreement between managers, the share of the variation due to idiosyncratic preferences by far outweighs that share of the variation explained by common preferences. Focusing on the agreement on the ranking as a manager, we estimate that around 20% (for the line management vignette) to 35% (for the supplier vignette) of the total variation in the latent utility is explained

⁹ We again estimate this using Hamiltonian Monte Carlo using the Stan language.

¹⁰ We refer to self-employment and wage-employment as two distinct types of employment throughout.

¹¹ This approach is broadly analogous to the identification strategy of [Bertrand and Schoar \(2003\)](#), which involves comparing R^2 values when adding manager fixed effects; this is conceptually similar to the role played by μ_i .

by the common component μ .

Table A.25: Dyadic agreement of managers across vignettes

	Vignette					Total
	Line Man- agement	Supplier	Pay rise	Bank	Client	
Ranking as manager						
Probability of agreement	56.7	60.9	58.0	57.7	59.3	58.5
Variance μ	0.26	0.78	0.41	0.48	0.60	0.58
Ranking as entrepreneur						
Probability of agreement	57.9	57.9	56.3	60.2	60.5	58.5
Variance μ	0.43	0.53	0.25	0.70	0.74	0.52

Notes This table depicts the probability of agreement and the variance of μ . The first is the probability that two managers assessing the same pair of candidates agree on their ranking as an entrepreneur and as a manager, split by the two types of employment and across vignettes. The variance of μ provides a measure of the homogeneity of preferences across managers. To decompose the heterogeneity in preferences for management traits across firms, we first analyse whether the dyadic agreement between firms can be predicted by observable characteristics. To do so, we implement median splits of the sample based on a number of observable characteristics. We then create three indicator variables: low_{ij} , equal to one if both firms are below median for that characteristic, $different_{ij}$ equal to one if one firm is below, and one firm above median, and $high_{ij}$ if both firms are above the median.

We then analyse the resulting data using the following regression model for pairs of firms $i \neq j$ with standard errors clustered at the pair level:

$$agree_{ij} = \alpha + \beta_1 different_{ij} + \beta_2 high_{ij} + \varepsilon_{ij} \quad (5)$$

Table A.26 reports the results from this regression. These results further stress that agreement between the HR managers is relatively low, and show the relationship with observable characteristics is limited. We see effectively no differences in rates of agreement by management score as calculated based on MOPS-type questions, number of employees, share of employees in a management position at the firm and share of non-payroll employees (in respectively columns 1, 2, 4 and 5). We only observe a clear increase in two characteristics: the number of competitors and managers' self-reported trust. Firms in more competitive environments appear to be more likely to agree (column 1), where the effect seems to be additive although only significant when both firms have a high number of competitors. The most clear relationships seems to be for high self-reported trust, an indicator for whether the manager reports to be generally trusting. We

find that the rate of agreement drops sharply when either manager reports to be trusting (column 6). This suggests that non-trusting managers pick up on some dimension of the responses, and trusting managers ignore this dimension without coordinating on some other observable instead. This is a sharp difference, with either manager being trusting decreasing the probability of agreement from 62.8% to around 55% in a binary choice.

Table A.26: Agreement between HR managers by firm and manager characteristics

Split by:	(1) Management Score	(2) Number of employees	(3) Number of competitors	(4) Share of employees in management	(5) Share of non-payroll employees	(6) Manager self- reported trust
One high, one low	0.00885 (0.0235)	0.0252 (0.0245)	0.0331 (0.0230)	-0.00837 (0.0261)	0.000646 (0.0216)	-0.0712 (0.0223)
Both high	0.00226 (0.0271)	0.0392 (0.0289)	0.0687* (0.0326)	0.0218 (0.0287)	-0.0166 (0.0317)	-0.0831** (0.0299)
Constant	0.580*** (0.0182)	0.563*** (0.0202)	0.558*** (0.0190)	0.584*** (0.0217)	0.587*** (0.0141)	0.628 *** (0.0163)
N	3948	3948	3849	3860	3948	3553

Notes This table reports the parameter estimates of regression 5. The column labels indicate the variables based on which the binary split of the sample for each regression is made. The dependent variable is the share of agreement between the two HR managers assessing the same pair of candidates. The variables on which the splits are made are the overall management score, the number of employees, the number of competitors, the share of employees in management positions, the share of non-payroll employees, and the self-reported trust of the manager. The number of observations in column (6) is lower due to some observations missing for this variable. The standard errors are in parentheses and the p-values are reported as: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

J Worker preferences over managers

This paper shows that firm managers strongly favor the rule-based managerial type. A natural follow-up is whether employees – who would work under such managers – hold similar preferences. To explore this, we carried out a supplementary data-collection exercise with workers supervised by entry-level managers, presenting them with the vignette recordings and gathering their evaluations through a series of closed- and open-ended questions.

Specifically, in May 2025 we revisited 17 firms from our sample.¹² At each firm, we asked the on-site

¹² After contacting the IRB at Oxford University to confirm our updated protocol was appropriate and covered by our existing ethics approval.

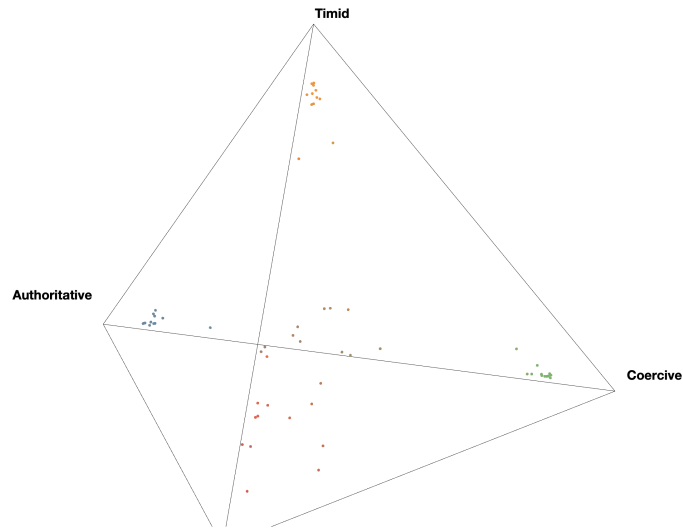


Figure A.13: Distribution of $\hat{\theta}_i$ for individuals in focus groups

manager to share the phone numbers of up to five of their employees in their main production process, selecting those who are managed by the lowest level of managers in the firm. We then contacted these workers and asked them to participate in a focus group in central Addis Ababa.

We ran these focus groups with groups of four to seven participants. We focused on the scenarios related to managing employees, so the line-management and pay-rise scenarios. We designed the exercise to be model-informed – after estimating the model – showing, for each of the two vignettes, recordings of four ‘archetypal’ managers and one ‘average’ manager. The “archetypal” managers are those most closely resembling each of the pure types in our sample. For example, for the rule-based type we shared the contact details of the most rule-based managers in our sample based on θ_i , specifically from the 95th percentile of the distribution. We then asked an enumerator to contact them going from most to least reliant on a rule-based managerial style within this subsample to ask for consent to show their videos to workers. We repeated this for each pure type. In practice, we always selected among the 20 managers closest to each pure type. For the “average” manager, we created this list by first calculating the entropy of classifying each individual based on θ_i and contacted the managers with the highest classification uncertainty. Figure A.13 shows the distribution of $\hat{\theta}_i$ for individuals included in the simplex, showing they are drawn from either the corners of the simplex – the pure types – and from the centroid of the simplex - the average types. Rule-based, power-based and discretion-based managers are all close to the corners of the simplex, whereas affiliative managers are further from their respective corner in terms of $\hat{\theta}_i$ as there are few “pure” affiliative managers.

Because workers might evaluate male and female managers differently for reasons unrelated to managerial style, we stratified the vignettes by manager gender. Given our modest sample size, each participant viewed recordings featuring exclusively male or exclusively female managers: the first nine focus groups

saw male managers and the final three saw female managers. This design mirrors the composition of our sample of young professionals, in which 78 percent of participants are male.

The average worker in our sample was 31.2 years old, had completed 12.8 years of schooling, had spent four years with the current firm and accumulated 7.5 years of labour-market experience (Table A.27). The sample was gender balanced (47.0 percent male). Most respondents worked in manufacturing (51.5 percent) or services (42.4 percent), with the remainder spread across smaller sectors. Their roles were concentrated in sales, administrative, and other customer-facing functions rather than in production.¹³ Additional summary statistics appear in Table A.28.

Table A.27: Summary Statistics – Continuous Variables

	mean	p50	sd	min	max
Age	31.2	30	9.7	19	57
Years of education	12.8	12	2.5	8	18
Experience with firm (years)	4.0	2	4.9	1	26
Labour market experience (years)	7.5	6	7.1	1	31
Observations	66				

Turning to the assessments. In the first part of this exercise, we asked the respondents to, separately for the two scenarios, watch five videos of young professionals responding to the scenario. After showing the respondents each video, we asked them individually to (a) list up to three words describing that person and (b) five Likert-style questions on a 5-point scale from Strongly Disagree to Strongly Agree (listed below):¹⁴

- (i). **Productivity** This person would raise the productivity of teams in my firm.
- (ii). **Targets** This person would help the firm achieve its targets.
- (iii). **Motivating** I would be happy to work hard for this person.
- (iv). **Consistent** This manager would treat different employees consistently.
- (v). **Typical** This type of response is typical for a manager in my firm.

After watching all five videos for one vignette, we asked the respondents to rank these five videos; to remind respondents which managers they had seen, we showed them photos of the five managers at this point. Enumerators reported that respondents were generally able to recall the managers based on these screenshots. We specifically asked the following question:

¹³ Weekday scheduling initially excluded production workers, who were indispensable to ongoing operations. After the first recruitment wave, most sessions were therefore moved to Saturdays, which are typically half-days in Ethiopia.

¹⁴ The label in bold text is reported in all regression tables as a shorthand.

Variable	Category	Count	Percent
Gender	1 = Male	31	47.0
Gender	2 = Female	35	53.0
Position	1. Intern / Trainee / Apprentic	1	1.5
Position	2. Administrative / Clerical Assistant	9	13.6
Position	3. Customer-facing Service / Teller	14	21.2
Position	4. Sales / Marketing Assistant	29	43.9
Position	5. Technical / IT Support Junior	7	10.6
Position	6. Production / Operator / Field Assistant	5	7.6
Position	7. Junior Professional (Associate / Analyst)	1	1.5
Sector	1 = Manufacturing	34	51.5
Sector	2 = Services	28	42.4
Sector	3 = Finance	2	3.0
Sector	4 = ICT	1	1.5
Sector	5 = Public/NGO	1	1.5
Education	Primary school	2	3.0
Education	High school	22	33.3
Education	Vocational school	7	10.6
Education	Diploma	10	15.2
Education	BA	21	31.8
Education	MSc	4	6.1
Enumerator	Kindalem	22	33.3
Enumerator	filsen	15	22.7
Enumerator	meskrem	29	43.9

Now, we would like you to rank these five responses from the most effective manager to the least effective manager; by effective we mean someone who would contribute to the long-term success of the firm he works for. We will show you photos of each of the aspiring managers you just saw and ask you to rank them.

We estimate two models using Stata to analyze responses: a rank-ordered logit model for rankings and an ordered logit model for Likert-scale ratings.

```
ologit rank i.type, group(respondentVignetteUniqueID) reverse
```

where `rank` is the rank from the respondent from one to five, `i.type` a dummy for being each “pure” type or the average type, and `respondentVignetteUniqueID` groups data at the respondent-vignette level. We report the estimates in ordered logit coefficient space.

For the Likert data:

```
ologit Y i.type , cluster(respondentUniqueID)
```

where Y_i is the score from the respondent, `i.type` a dummy for being each “pure” type or the average type, and `respondentUniqueID` a respondent ID. We currently report results in logit space.

Table A.29 provides the main results from this exercise, where the discretion-based type is omitted. Respondents show a clear preference for a rule-based management style among the five responses. This pattern is reflected in both the ranking data in column 1, as well as for each individual Likert question. When we look at the remaining types, we see both the power-based and Average type are perceived as better than the discretion-based type, and largely for very similar reasons. Interestingly, the affiliative type is perceived as the worst type.

Table A.29: Worker's perceptions of management styles

	(1) Ranking	(2) Productivity	(3) Targets	(4) Motivating	(5) Consistent	(6) Typical
main						
Rule-based	0.938*** (0.163)	1.292*** (0.241)	1.468*** (0.257)	1.039*** (0.240)	1.286*** (0.226)	0.720*** (0.213)
Affiliative	-0.269* (0.158)	-0.801*** (0.223)	-0.595** (0.244)	-0.649*** (0.220)	-0.652*** (0.209)	-0.714*** (0.215)
Power-based	0.387** (0.156)	0.247 (0.243)	0.526** (0.229)	0.208 (0.198)	0.218 (0.240)	0.277 (0.217)
Average	0.429*** (0.157)	0.308 (0.196)	0.383* (0.225)	0.460** (0.194)	0.416* (0.213)	0.308 (0.201)
/						
cut1		-2.832*** (0.232)	-3.177*** (0.272)	-2.121*** (0.174)	-2.666*** (0.197)	-2.314*** (0.202)
cut2		-0.985*** (0.178)	-0.986*** (0.200)	-0.823*** (0.162)	-0.997*** (0.184)	-0.581*** (0.170)
cut3		-0.253 (0.194)	-0.194 (0.200)	-0.154 (0.161)	-0.146 (0.174)	0.143 (0.154)
cut4		1.726*** (0.231)	2.052*** (0.238)	1.925*** (0.210)	2.017*** (0.212)	2.520*** (0.232)
Rule-based=Affiliative	0.000	0.000	0.000	0.000	0.000	0.000
Rule-based=Power-based	0.001	0.000	0.001	0.002	0.000	0.068
Rule-based=Average	0.001	0.000	0.000	0.015	0.000	0.074

Notes: Standard errors clustered by respondent. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finally, we implement this for the two vignettes separately. We report the results in Tables A.30 and A.31 for the line management and pay rise scenario respectively. We find broadly similar trends across the two scenarios, although the discretion-based type appears to perform relatively better in the line management scenario than in the pay rise scenario when compared to the other types.

Table A.30: Line Management: Worker’s perceptions of management styles

	(1) Productivity	(2) Targets	(3) Motivating	(4) Consistent	(5) Typical	(6) Ranking
main						
Rule-based	0.684*** (0.231)	1.397*** (0.334)	1.753*** (0.363)	0.922*** (0.328)	1.629*** (0.334)	0.626** (0.275)
Affiliative	-0.529** (0.228)	-0.825*** (0.291)	-0.577* (0.326)	-0.767** (0.308)	-0.546* (0.297)	-0.568* (0.326)
Power-based	0.278 (0.230)	0.498 (0.338)	0.897*** (0.319)	0.205 (0.305)	0.581 (0.357)	0.373 (0.343)
Average	0.211 (0.222)	0.202 (0.306)	0.261 (0.344)	0.240 (0.309)	0.588* (0.308)	0.266 (0.338)
/						
cut1		-2.753*** (0.287)	-3.195*** (0.415)	-2.320*** (0.253)	-2.742*** (0.316)	-2.450*** (0.243)
cut2		-0.971*** (0.222)	-0.771*** (0.230)	-0.966*** (0.207)	-0.859*** (0.236)	-0.555** (0.226)
cut3		-0.254 (0.232)	-0.0982 (0.235)	-0.359* (0.215)	-0.109 (0.227)	0.156 (0.230)
cut4		1.850*** (0.292)	2.216*** (0.291)	1.859*** (0.261)	2.336*** (0.281)	2.577*** (0.320)
Observations	305	305	305	305	305	305

Notes: Using only data from the line management scenario. Standard errors clustered by respondent. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.31: Pay Rise: Worker's perceptions of management styles

	(1) Productivity	(2) Targets	(3) Motivating	(4) Consistent	(5) Typical	(6) Ranking
main						
Rule-based	1.184*** (0.232)	1.206*** (0.312)	1.205*** (0.319)	1.151*** (0.281)	1.021*** (0.298)	0.821** (0.324)
Affiliative	-0.0287 (0.220)	-0.776** (0.323)	-0.628* (0.333)	-0.546* (0.323)	-0.749*** (0.290)	-0.856*** (0.295)
Power-based	0.491** (0.215)	0.0239 (0.342)	0.179 (0.340)	0.221 (0.271)	-0.0845 (0.302)	0.195 (0.307)
Average	0.641*** (0.223)	0.387 (0.264)	0.464 (0.301)	0.655** (0.272)	0.270 (0.311)	0.342 (0.288)
/						
cut1		-2.917*** (0.339)	-3.192*** (0.358)	-1.947*** (0.268)	-2.626*** (0.276)	-2.209*** (0.302)
cut2		-1.000*** (0.260)	-1.209*** (0.291)	-0.692*** (0.229)	-1.120*** (0.244)	-0.607** (0.246)
cut3		-0.250 (0.263)	-0.295 (0.269)	0.0339 (0.214)	-0.175 (0.236)	0.132 (0.225)
cut4		1.626*** (0.262)	1.919*** (0.300)	1.991*** (0.251)	1.764*** (0.262)	2.475*** (0.290)
Observations	329	329	329	329	329	329

Notes: Using only data for the pay rise scenario. Standard errors clustered by respondent. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

K Summary statistics

Table A.32: Descriptive statistics of young professionals

Categorical Variables				
Gender (% Male)	78%			
University degree	78%			
Self-employed	17%			
Wage-employed	68%			
Neither parent completed university	88%			
Neither parent completed primary school	48%			
Conditional on Wage Employment				
In a professional position	98%			
In a managerial position	18%			
Continuous Variables				
	Median	Std. Dev.	5th Perc.	95th Perc.
Age (years)	31	2.8	27	36
Wage (ETB/Month)	6000	5979	2285	18000
Profit (ETB/Month)	6000	16099	0	42000
<i>N</i>	982			

Notes This table provides summary statistics of the respondents included in the sample. The wage and profit in Ethiopian Birr (ETB) are calculated conditional on being in wage- and self-employment, respectively. The percentages of individuals in professional and managerial positions are conditional on being in wage employment.

Table A.33: Descriptive statistics of firm managers

Categorical Variables				
BA Degree		79.7%		
MSc Degree		24.0%		
Formal management education		72.4%		
Department in firm				
Finance		12.6%		
HR		40.5%		
Administration		34.6%		
Other		12.3%		
Continuous Variables				
	Median	Std. Dev.	5th Perc.	95th Perc.
Years of experience at this firm	8	8.02	1	27
Years of labour market experience	16	14.56	4	46
<i>N</i>	587			

Notes This table provides summary statistics of the HR managers included in the sample. 11 firms were omitted from the main analysis as they ranked triplets of candidates that had already been ranked by two other firms.

Table A.34: Descriptive statistics of firms

Sector	Count	Share
Construction	51	8.69%
Education	27	4.60%
Healthcare	30	5.11%
Manufacturing	173	29.47%
Services	167	28.45%
Trade	86	14.65%
Other	53	9.03%

Continuous Variables	Median	Std. Dev.	5th Perc.	95th Perc.
Number of payroll employees	58	1771.478	6	700
Share of female payroll employees	40.0%	19.2%	11.4%	73.1%
<i>N</i>	587			

Notes This table provides summary statistics of the firms included in the sample. 11 out of these 587 firms were omitted from the main analysis as they ranked triplets of candidates that had already been ranked by two other firms.

L Robustness Gender Tables

In this section, we report the multinomial logit results of regressions of the gender of the actor seen in the first vignette on the attribute exhibited by the respondent in the subsection vignettes.

Table A.35: Gender Robustness: Agreement

	Agree b/se
First actor female	-0.120 (0.097)
Constant	-1.665*** (0.070)
Observations	6887

Notes: Multinomial logit coefficients with standard errors in parentheses. The dependent variable indicates whether the respondent agreed with the action in the vignette. Statistical significance is denoted by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.36: Gender Robustness: Source of Authority

	Source of authority b/se
<i>A formal policy</i>	
First actor female	0.179 (0.112)
Constant	-1.246*** (0.083)
<i>Higher principles</i>	
First actor female	0.417 (0.312)
Constant	-3.821*** (0.242)
<i>Personal relationship</i>	
First actor female	0.678 (0.429)
Constant	-4.626*** (0.340)
<i>No source of authority</i>	
First actor female	0.026 (0.093)
Constant	-0.605*** (0.067)
Observations	6887

Notes: Multinomial logit coefficients with standard errors in parentheses. The omitted category is reference to seniority. Statistical significance is denoted by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.37: Gender Robustness: Justification

	Justification b/se
<i>Other party's interest</i>	
First actor female	-0.123 (0.164)
Constant	-2.428*** (0.118)
<i>Action is in respondent's interest</i>	
First actor female	-0.225* (0.107)
Constant	-1.134*** (0.075)
<i>Reference to relationship</i>	
First actor female	-0.150 (0.116)
Constant	-1.630*** (0.082)
<i>No justification given</i>	
First actor female	0.004 (0.128)
Constant	-1.604*** (0.096)
Observations	6887

Notes: Multinomial logit coefficients with standard errors in parentheses. The omitted category is reference to protecting/helping the company. Statistical significance is denoted by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.38: Gender Robustness: Tone

	Tone b/se
<i>Aggressive</i>	
First actor female	-0.079 (0.168)
Constant	-2.262*** (0.119)
<i>Assertive</i>	
First actor female	-0.014 (0.089)
Constant	-0.401*** (0.063)
Observations	6887

Notes: Multinomial logit coefficients with standard errors in parentheses. The omitted category is a calm/assured tone. Statistical significance is denoted by * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

M Text analysis

M.1 Text cleaning

We first clean this text in the following steps:

- (i). We make all text lower case
- (ii). We remove all punctuation.
- (iii). We remove common stop words in Python using the basic lists from the packages `nltk_sw`, `wc_sw` and `spacy_sw`.

For the analyses where this is relevant, we also drop all words which occur in more than 90% of the documents, and drop all words which occur fewer than three times in the full corpus. We explicitly state this when we do so in each subsection. We also apply row-wise L2-normalization to document vectors where appropriate so results are not driven by response length.

M.2 Word clouds

To better understand the content of young professionals' responses, we first generate simple word clouds based on term frequency for each individual vignette. We primarily do this to verify the broad content of the vignette responses is fitting for the scenarios.



Figure A.14: Line management scenario



Figure A.15: Supplier scenario



Figure A.18: Client scenario

M.3 Text features predicting types

To better understand the language patterns associated with our four managerial styles, we study correlations between $\hat{\theta}$ and the words used by individuals in their responses. As types are determined at the individual level, we pool the data from all five vignettes for this exercise. Our objective is to determine what words correlate with higher weight on each managerial style, which we visualize in word clouds.

We began by constructing a document–term matrix from the set of translations. Words were lowercased, and we kept only those that appeared in at least three documents but not in more than ninety percent of documents. To remove uninformative words, we applied a broad set of stopword lists as described above, and also dropped very short tokens and purely numeric entries.

Each document was matched to the individuals’ managerial type weights $\hat{\theta}$. We then carried out five-fold cross-validation grouped at the individual level: in each fold, we computed the correlation between every word and each outcome, while restricting attention to words that appeared in that fold’s training set.

To identify reliable predictors, we used a simple stability-selection rule. Words are retained if they appear in at least three folds and their correlation with the outcome had a consistent sign at least eighty percent of the time. Among those retained, we ranked words by the strength of their average correlation and selected up to the top 60 positive and 60 negative predictors for each outcome.

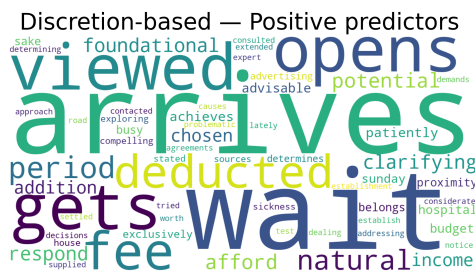
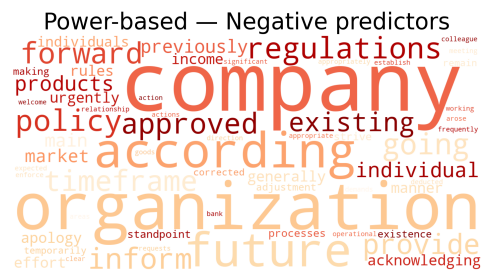
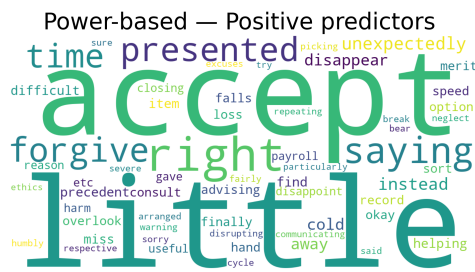
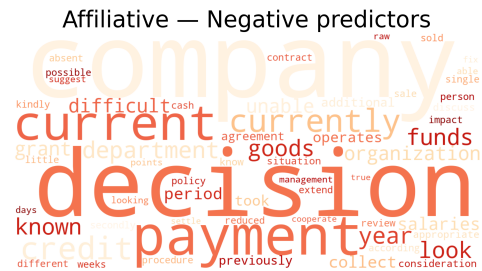
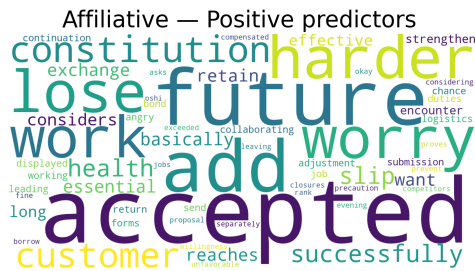
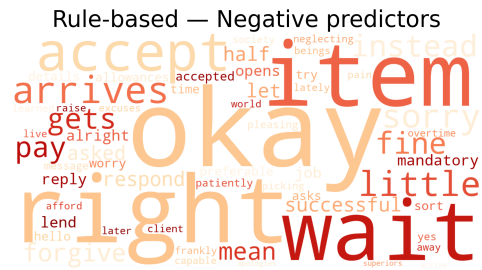
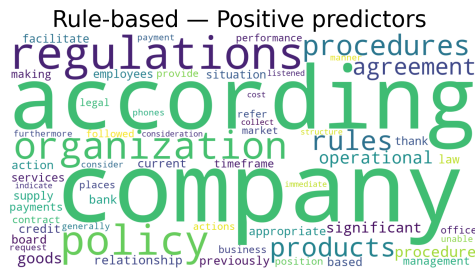


Figure A.20: Positive and negative words

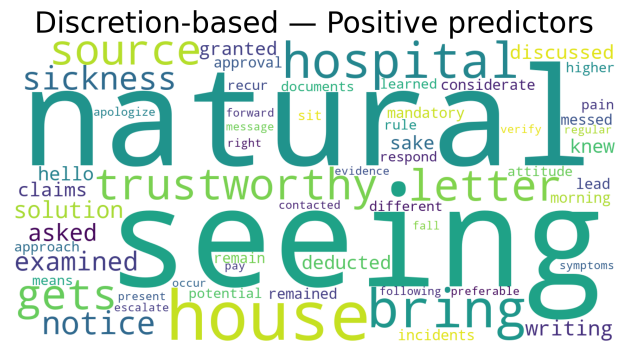
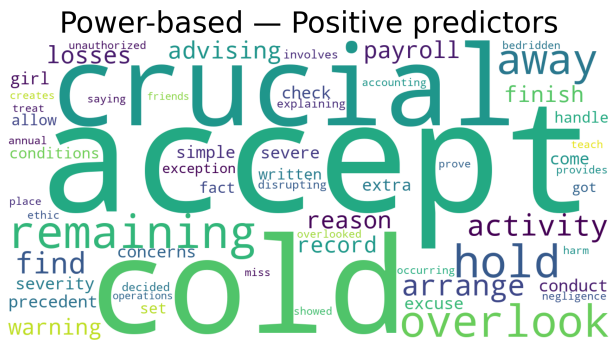
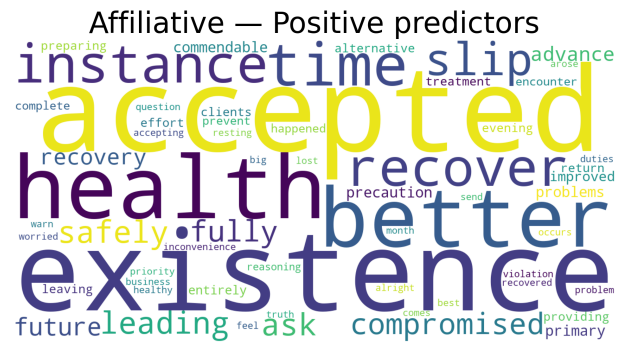
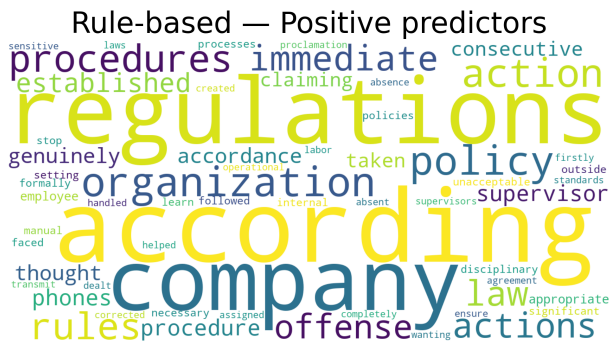


Figure A.21: Line management: Positive predictors

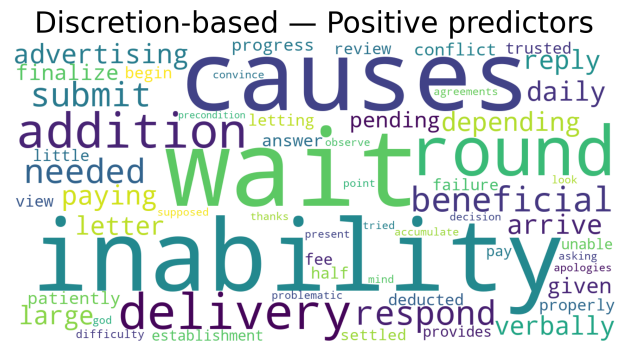
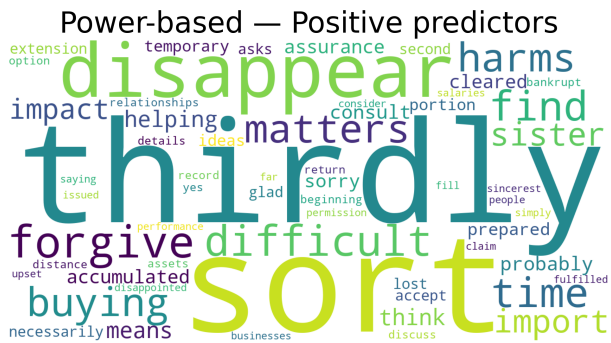
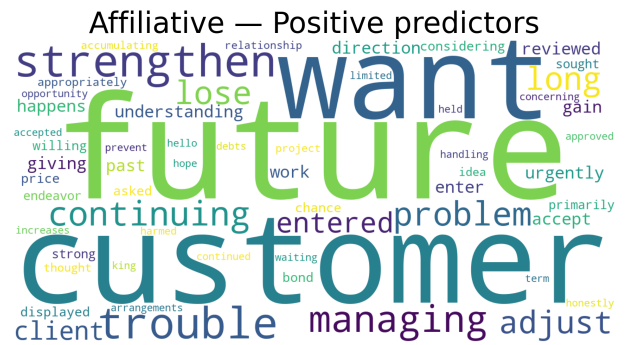


Figure A.25: Client: Positive predictors

M.4 Semantic word clusters and managerial styles

In addition to examining individual words, we also seek to capture broader semantic patterns by clustering words according to their meanings. To represent meaning, we rely on a pre-trained embedding model (spaCy’s large English vectors), which maps each word into a high-dimensional vector space where semantically similar words are located near one another.

We then filter the vocabulary to include only words that appear in at least three documents (to ensure stability) but in less than 90% of documents (to avoid trivial, overly common terms). The remaining words are grouped into seventy clusters using k-means, an algorithm that partitions the embedding space into regions of similar words (on L2-normalized vectors). Each cluster is summarized by its “prototypical words,” namely those that lie closest to the cluster centroid, which makes interpretation straightforward.

We choose to implement this methodology with 70 semantic word clusters. This choice is based on a combination of the elbow method and the silhouette score, both computed in scikit-learn. Figure A.26 reports results for K from 10 to 210. The left panel plots cosine-inertia (one minus cosine to the assigned centroid) and shows no clear elbow; applying a standard “maximum distance to chord” heuristic to the *log*-inertia curve – i.e., the maximum perpendicular distance to the straight line joining the endpoints – suggests 70 clusters (Satopää, Albrecht, Irwin, and Raghavan, 2011). The right panel plots the average silhouette score using cosine distance, which flattens at K=70 (Rousseeuw, 1987). We choose to set K = 70 in line with the suggestion of the elbow plot, balancing fit and interpretability. We note results are qualitatively similar for higher numbers of clusters.

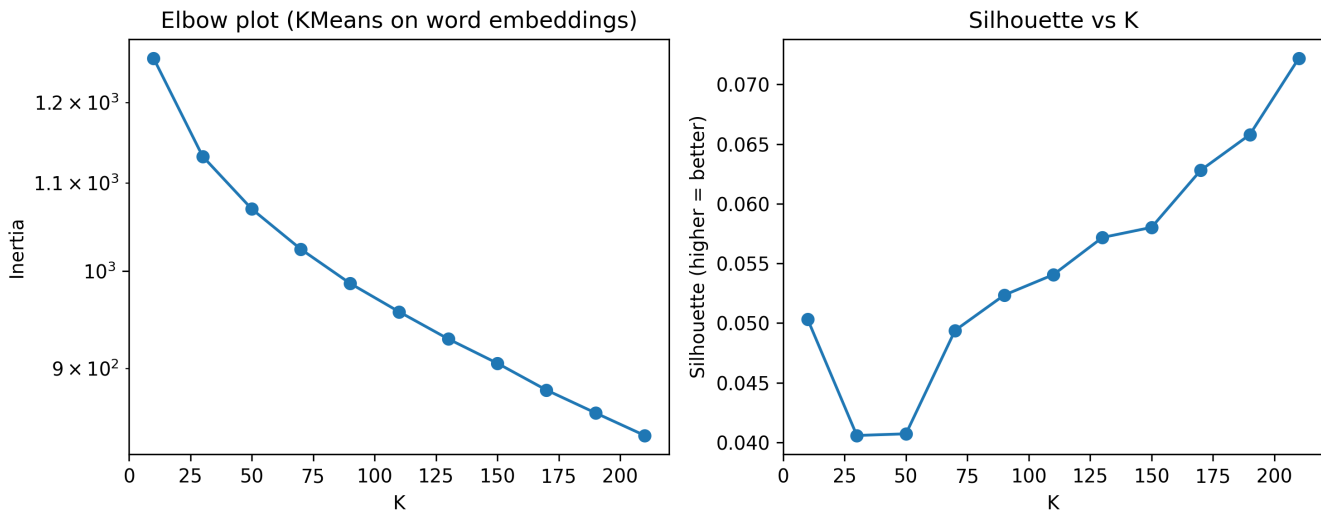


Figure A.26: Choice of the number of clusters

At the document level, we quantify how much each text draws on each semantic cluster. We implement a “soft” measure, which accounts for graded similarity by weighting words according to their closeness to

the cluster centroid.¹⁵ We truncate negative cosine similarities at zero, since negative values are difficult to interpret semantically. To avoid confounding with document length, these cluster-level features are normalized so that each document’s profile reflects relative emphasis rather than absolute word counts (by dividing by the sum of soft cluster weights per document).

We then relate the cluster profiles to the four managerial traits. First, we compute simple correlations to see which clusters tend to increase or decrease with each element of $\hat{\theta}$. Second, we fit a regression model that relates the semantic clusters to the four outcome dimensions. Because clusters can overlap in meaning and are often correlated with each other, a standard regression would risk overfitting and producing unstable estimates. To address this, we use an adapted ridge regression which we describe after this paragraph. This method applies an L_2 penalty that shrinks coefficients toward zero when the evidence is weak or when predictors are highly collinear. This penalization reduces noise, improves stability across clusters, and helps ensure that the words or clusters we highlight are genuinely robust signals rather than statistical artefacts. Finally, for each outcome we highlight the clusters most strongly predictive of higher or lower values, together with their most representative words. This analysis moves beyond single-word predictors and provides a richer picture of how semantically related groups of terms map onto the latent behavioural types.

Specifically, instead of a standard ridge regression we train a fractional logit regression with ℓ_2 penalty (inverse strength $C = 1.0$) *directly on the soft targets* by minimizing cross-entropy,

$$\mathcal{L}(\beta) = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K \hat{\theta}_{ik} \log P_{ik} + \lambda \sum_{k=1}^K \|\beta_k\|_2^2, \quad P_{ik} = \frac{\exp(X_i^\top \beta_k)}{\sum_{\ell=1}^K \exp(X_i^\top \beta_\ell)}. \quad (6)$$

This objective equals $\frac{1}{N} \sum_i \text{KL}(\hat{\theta}_i \| P_i)$ up to a constant, so no separate calibration step is used and final predictions $P_i = (P_{i1}, \dots, P_{iK})$ are the model probabilities.¹⁶

In the following tables, we report the 10 clusters with the strongest positive correlation with each managerial style, together with the “ridge” coefficient from the multinomial logistic regression and the six words closest to the center of each cluster (prototypical words).

¹⁵ This implies that, for example, the word “power” is related to both “authority” and “energy”.

¹⁶ Implemented in scikit-learn by weighting class-specific copies so that the weighted log-loss matches cross-entropy with soft targets.

M.5 Clusters with the strongest positive correlation with each type

Table A.39: Top semantic clusters associated with a rule-based managerial style ($\hat{\theta}_1$)

Corr.	Ridge	Prototypical words
0.231	+0.003	institution (216), policy (210), country (147), public (72), state (69), government (68)
0.200	+0.012	colleagues (111), cooperate (72), cooperation (59), collaboration (16), collaborative (9), collaboratively (7)
0.181	-0.039	organization (1665), currently (377), board (116), established (80), served (58), behalf (45)
0.178	-0.150	report (90), reviewed (81), study (60), reviewing (60), effectiveness (43), guidelines (32)
0.174	+0.240	office (303), management (250), supervisor (187), manager (183), responsibility (112), tasks (97)
0.174	-0.192	effort (118), efforts (61), community (59), successful (54), resources (52), organizations (46)
0.173	+0.190	regulations (279), rules (209), law (126), legal (113), standards (99), penalty (70)
0.153	+0.194	presented (651), understanding (137), perspective (113), discussed (90), presenting (37), demonstrate (25)
0.146	+0.175	system (111), internal (55), control (41), workstation (30), storage (17), safety (16)
0.139	-0.106	benefits (115), benefit (101), achieve (92), contribution (83), achieved (82), contributions (51)

Notes: Corr. is the Pearson correlation between the L1-normalized soft cluster profile and $\hat{\theta}_1$. Ridge reports coefficients from ridge regression (Hoerl and Kennard, 1970). Soft memberships use positive cosine similarity only (negatives truncated to zero); profiles are L1-normalized per document. Prototypical words are nearest to each cluster centroid in embedding space.

Table A.40: Top semantic clusters associated with an affiliative managerial style ($\hat{\theta}_2$)

Corr.	Ridge	Prototypical words
0.086	+0.016	future (876), bring (738), continue (603), need (587), able (399), advance (217)
0.082	-0.105	know (1564), good (1277), want (1263), reason (1213), better (778), things (624)
0.079	+0.084	thank (1519), hello (338), dear (223), appreciate (186), happy (179), hope (169)
0.074	+0.110	warning (307), unable (285), lose (201), cause (112), bear (106), mistake (94)
0.072	+0.105	okay (1997), sick (771), yes (360), sorry (252), alright (114), excuse (92)
0.071	+0.200	difficult (350), quickly (213), bit (185), clear (78), completely (55), thoroughly (52)
0.069	-0.020	come (801), long (408), right (397), wait (347), leave (334), little (324)
0.066	-0.110	diligent (269), hardworking (112), loyal (69), confident (26), assured (25), honest (21)
0.064	+0.113	trust (171), diligence (84), desire (56), commitment (54), grace (53), patience (52)
0.057	-0.052	family (152), mother (88), friends (84), home (81), brother (74), friend (71)

Notes: Corr. is the Pearson correlation between the L1-normalized soft cluster profile and $\hat{\theta}_2$. Ridge reports coefficients from ridge regression (Hoerl and Kennard, 1970). Soft memberships use positive cosine similarity only (negatives truncated to zero); profiles are L1-normalized per document. Prototypical words are nearest to each cluster centroid in embedding space.

Table A.41: Top semantic clusters associated with a power-based managerial style ($\hat{\theta}_3$)

Corr.	Ridge	Prototypical words
0.142	+0.032	know (1564), good (1277), want (1263), reason (1213), better (778), things (624)
0.135	-0.127	come (801), long (408), right (397), wait (347), leave (334), little (324)
0.127	-0.305	okay (1997), sick (771), yes (360), sorry (252), alright (114), excuse (92)
0.120	+0.018	difficult (350), quickly (213), bit (185), clear (78), completely (55), thoroughly (52)
0.120	+0.101	warning (307), unable (285), lose (201), cause (112), bear (106), mistake (94)
0.118	+0.032	thank (1519), hello (338), dear (223), appreciate (186), happy (179), hope (169)
0.117	+0.181	family (152), mother (88), friends (84), home (81), brother (74), friend (71)
0.114	-0.039	moment (89), patiently (88), listened (44), reminding (39), listen (33), forgotten (25)
0.114	+0.173	took (581), said (517), told (498), mentioned (330), heard (308), called (287)
0.104	+0.078	question (547), note (224), answer (151), explain (146), written (122), discussing (93)

Notes: Corr. is the Pearson correlation between the L1-normalized soft cluster profile and $\hat{\theta}_3$. Ridge reports coefficients from ridge regression (Hoerl and Kennard, 1970). Soft memberships use positive cosine similarity only (negatives truncated to zero); profiles are L1-normalized per document. Prototypical words are nearest to each cluster centroid in embedding space.

Table A.42: Top semantic clusters associated with a discretion-based managerial style ($\hat{\theta}_4$)

Corr.	Ridge	Prototypical words
0.067	-0.055	time (2837), days (860), year (680), day (348), weeks (219), years (168)
0.053	-0.090	second (592), return (416), short (230), half (147), end (94), lead (92)
0.053	+0.017	today (231), coming (220), soon (198), arrive (131), expected (110), earlier (94)
0.050	-0.103	collect (40), unsold (30), collected (15), accumulating (12), collecting (10), piled (10)
0.046	+0.190	pay (2138), money (975), payment (960), paid (421), paying (226), cash (135)
0.044	+0.212	deadline (164), notify (113), delay (109), date (108), delayed (88), terminate (84)
0.032	+0.106	items (695), sold (254), sell (253), price (223), stock (153), store (106)
0.030	+0.012	previously (257), received (243), taken (176), returned (145), initially (123), raised (110)
0.030	-0.029	cold (151), main (144), cover (121), accommodate (98), small (85), single (57)
0.029	+0.243	come (801), long (408), right (397), wait (347), leave (334), little (324)

Notes: Corr. is the Pearson correlation between the L1-normalized soft cluster profile and $\hat{\theta}_4$. Ridge reports coefficients from ridge regression (Hoerl and Kennard, 1970). Soft memberships use positive cosine similarity only (negatives truncated to zero); profiles are L1-normalized per document. Prototypical words are nearest to each cluster centroid in embedding space.

M.6 Person-Level Text Classifier

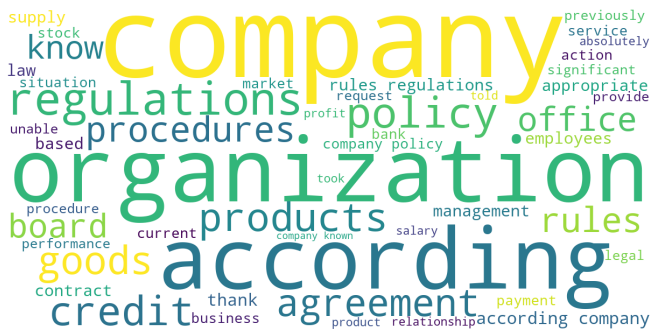
To allow for more semantic meaning than word clouds, but in a simpler framework than the semantic cluster analysis, we build on the analysis in section M.3 to predict type membership using tokens (unigrams and bigrams) instead of single words. For simplicity, for each individual $i \in \{1, \dots, N\}$, we concatenate all of their free-text responses into a single document D_i . Next, we featurize D_i with a TF-IDF vectorizer on word unigrams and bigrams to obtain $X_i \in \mathbb{R}^p$, we first remove stopwords and common terms to focus on more semantically distinct terms, but explicitly keep all negations.

We then train a multinomial logistic regression with ℓ_2 penalty (inverse strength $C = 1.0$) *directly on the soft targets* by minimizing cross-entropy,

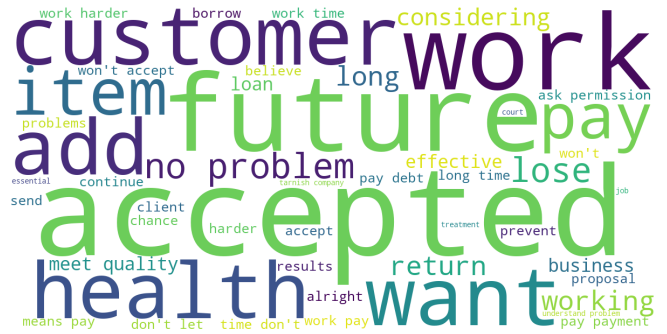
$$\mathcal{L}(\beta) = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K \hat{\theta}_{ik} \log P_{ik} + \lambda \sum_{k=1}^K \|\beta_k\|_2^2, \quad P_{ik} = \frac{\exp(X_i^\top \beta_k)}{\sum_{\ell=1}^K \exp(X_i^\top \beta_\ell)}.$$

This objective equals $\frac{1}{N} \sum_i \text{KL}(\hat{\theta}_i \| P_i)$ up to a constant, so no separate calibration step is used and final predictions $P_i = (P_{i1}, \dots, P_{iK})$ are the model probabilities.¹⁷

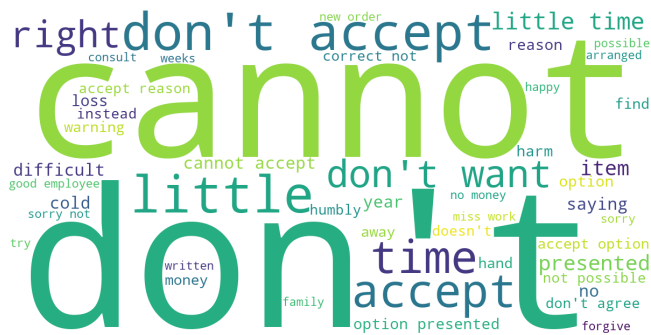
¹⁷ Implemented in scikit-learn by weighting class-specific copies so that the weighted log-loss matches cross-entropy with soft targets.



(a) most predictive tokens for rule-based



(b) most predictive tokens for affiliative



(c) most predictive tokens for power-based



(d) most predictive tokens for discretion-based

Figure A.27: Words best predicting type membership

N AI Analysis

To supplement our enumerator transcription, we conducted a similar transcription survey using OpenAI’s GPT-5 model.¹⁸ This allowed us to answer a number of questions complementary to our human enumerators transcriptions using only the text responses, which were also used in M.

To do this, we created a list of all transcriptions in Amharic labelled by the individual ID and vignette they related to. We then used OpenAI’s API to conduct the survey by uploading, for each transcription, both the questionnaire and the individual transcript, and downloading the resulting .json file. This procedure generated a structured dataset of AI-coded survey responses, fully analogous to the human-coded survey data used in the main analysis.

To appropriately prompt the model, the survey started with the following prompt:

You are an expert enumerator answering questions about hypothetical management tasks.
Give only numeric responses to closed-form questions. Output valid JSON.

N.1 Results from the AI Survey

We begin by presenting the results from the closed-form question. Where relevant, we construct binary indicators from the multinomial outcomes. Specifically, we focus on the following outcomes:

- Show empathy: Does the respondent show empathy to the person they are addressing?
- Show respect: Is the respondent respectful?
- Express distrust: Does the respondent express distrust?
- Mention future relationship: Does the respondent mention the future relationship between the two parties?
- Clear decision: Does the respondent make a clear decision?
- Strict policy: Does the respondent implement a strict policy?
- Address formally: Does the respondent use formal language to address the other party?
- Mention CV: Does the respondent refer to their CV?

Using these indicators, we replicate the analysis in Section 3, linking individuals’ observable characteristics to their type membership. Each individual is assigned to the pure type corresponding to their highest estimated $\hat{\theta}_i$, and we then compute conditional averages for each pure type. To test whether intentions

¹⁸ gpt-5-2025-08-07

differ across types, we compute the Mahalanobis (Wald) statistic and compare it to a χ^2_3 distribution, relying on the multivariate normal approximation to the posterior described in Section 4.1 of Gelman et al. (2013). For non-binary outcomes, we split the sample at the median to create a binary measure.

We first pool responses across the five vignettes (Table A.43), and then replicate the analysis separately by vignette (Tables A.44–A.48).

Throughout, the results are in line with our description of our main results. Interestingly, the rule-based type is most likely to show empathy, express distrust, and rely on procedural fairness to stress why this is a fair decision.

Table A.43: **Characteristics and Types: Summary Statistics**

Panel A: Characteristics related to interpersonal relationship				
Type	Show empathy	Show respect	Expresses distrust	Mentions future relation
Type 1	0.476	0.934	0.266	0.613
Type 2	0.466	0.951	0.205	0.561
Type 3	0.430	0.917	0.293	0.525
Type 4	0.404	0.938	0.223	0.481
<i>p</i> -value	0.000	0.003	0.000	0.000
Panel B: Characteristics related to contents of response				
Type	Clear decision	Strict policy	Address formally	Mentions CV
Type 1	0.846	0.283	0.212	0.002
Type 2	0.852	0.169	0.153	0.001
Type 3	0.845	0.195	0.169	0.002
Type 4	0.824	0.204	0.178	0.001
<i>p</i> -value	0.084	0.000	0.003	0.441

Notes These numbers are calculated by assigning each individual to the pure type for which they have the highest estimated $\hat{\theta}_i$. Then, a conditional average is taken for each pure type. To test whether intentions differ across types, we compute the Mahalanobis (Wald) statistic and compare it to a χ^2_3 distribution. This relies on the multivariate normal approximation to the posterior described in Section 4.1 of Gelman et al. (2013). For non-binary outcomes, we first create a binary split at the median to split the sample.

Table A.44: **Vignette 1: Characteristics and Types: Summary Statistics**

Panel A: Characteristics related to interpersonal relationship				
Type	Show empathy	Show respect	Expresses distrust	Mentions future relation
Type 1	0.367	0.825	0.557	0.443
Type 2	0.488	0.865	0.374	0.458
Type 3	0.375	0.773	0.594	0.410
Type 4	0.394	0.853	0.472	0.412
<i>p</i> -value	0.035	0.007	0.000	0.783

Panel B: Characteristics related to contents of response				
Type	Clear decision	Strict policy	Address formally	Mentions CV
Type 1	0.824	0.510	0.015	0.000
Type 2	0.816	0.369	0.008	0.000
Type 3	0.848	0.425	0.013	0.000
Type 4	0.798	0.405	0.012	0.000
<i>p</i> -value	0.245	0.000	0.864	-1.000

Notes These numbers are calculated by assigning each individual to the pure type for which they have the highest estimated $\hat{\theta}_i$. Then, a conditional average is taken for each pure type. To test whether intentions differ across types, we compute the Mahalanobis (Wald) statistic and compare it to a χ^2_3 distribution. This relies on the multivariate normal approximation to the posterior described in Section 4.1 of [Gelman et al. \(2013\)](#). For non-binary outcomes, we first create a binary split at the median to split the sample.

Table A.45: **Vignette 2: Characteristics and Types: Summary Statistics**

Panel A: Characteristics related to interpersonal relationship				
Type	Show empathy	Show respect	Expresses distrust	Mentions future relation
Type 1	0.402	0.935	0.336	0.799
Type 2	0.315	0.949	0.291	0.706
Type 3	0.356	0.931	0.320	0.713
Type 4	0.334	0.943	0.248	0.654
<i>p</i> -value	0.327	0.732	0.105	0.000

Panel B: Characteristics related to contents of response				
Type	Clear decision	Strict policy	Address formally	Mentions CV
Type 1	0.957	0.441	0.210	0.004
Type 2	0.947	0.272	0.145	0.004
Type 3	0.959	0.313	0.141	0.008
Type 4	0.944	0.308	0.181	0.000
<i>p</i> -value	0.910	0.000	0.019	0.471

Notes These numbers are calculated by assigning each individual to the pure type for which they have the highest estimated $\hat{\theta}_i$. Then, a conditional average is taken for each pure type. To test whether intentions differ across types, we compute the Mahalanobis (Wald) statistic and compare it to a χ^2_3 distribution. This relies on the multivariate normal approximation to the posterior described in Section 4.1 of [Gelman et al. \(2013\)](#). For non-binary outcomes, we first create a binary split at the median to split the sample.

Table A.46: **Vignette 3: Characteristics and Types: Summary Statistics**

Panel A: Characteristics related to interpersonal relationship				
Type	Show empathy	Show respect	Expresses distrust	Mentions future relation
Type 1	0.574	0.955	0.224	0.678
Type 2	0.564	0.970	0.171	0.630
Type 3	0.520	0.941	0.272	0.573
Type 4	0.446	0.955	0.196	0.547
<i>p</i> -value	0.001	0.378	0.021	0.000

Panel B: Characteristics related to contents of response				
Type	Clear decision	Strict policy	Address formally	Mentions CV
Type 1	0.949	0.219	0.177	0.001
Type 2	0.939	0.079	0.099	0.000
Type 3	0.923	0.114	0.110	0.001
Type 4	0.904	0.143	0.150	0.002
<i>p</i> -value	0.193	0.000	0.018	0.775

Notes These numbers are calculated by assigning each individual to the pure type for which they have the highest estimated $\hat{\theta}_i$. Then, a conditional average is taken for each pure type. To test whether intentions differ across types, we compute the Mahalanobis (Wald) statistic and compare it to a χ^2_3 distribution. This relies on the multivariate normal approximation to the posterior described in Section 4.1 of [Gelman et al. \(2013\)](#). For non-binary outcomes, we first create a binary split at the median to split the sample.

Table A.47: **Vignette 4: Characteristics and Types: Summary Statistics**

Panel A: Characteristics related to interpersonal relationship				
Type	Show empathy	Show respect	Expresses distrust	Mentions future relation
Type 1	0.449	0.997	0.003	0.471
Type 2	0.394	0.993	0.005	0.353
Type 3	0.382	0.995	0.007	0.338
Type 4	0.388	0.989	0.002	0.249
<i>p</i> -value	0.054	0.417	0.781	0.000

Panel B: Characteristics related to contents of response				
Type	Clear decision	Strict policy	Address formally	Mentions CV
Type 1	0.567	0.026	0.483	0.007
Type 2	0.619	0.042	0.414	0.001
Type 3	0.559	0.020	0.475	0.002
Type 4	0.564	0.027	0.408	0.001
<i>p</i> -value	0.598	0.232	0.077	0.159

Notes These numbers are calculated by assigning each individual to the pure type for which they have the highest estimated $\hat{\theta}_i$. Then, a conditional average is taken for each pure type. To test whether intentions differ across types, we compute the Mahalanobis (Wald) statistic and compare it to a χ^2_3 distribution. This relies on the multivariate normal approximation to the posterior described in Section 4.1 of [Gelman et al. \(2013\)](#). For non-binary outcomes, we first create a binary split at the median to split the sample.

Table A.48: **Vignette 5: Characteristics and Types: Summary Statistics**

Panel A: Characteristics related to interpersonal relationship				
Type	Show empathy	Show respect	Expresses distrust	Mentions future relation
Type 1	0.590	0.958	0.211	0.676
Type 2	0.575	0.977	0.187	0.658
Type 3	0.514	0.945	0.272	0.589
Type 4	0.457	0.948	0.197	0.542
<i>p</i> -value	0.000	0.321	0.022	0.000

Panel B: Characteristics related to contents of response				
Type	Clear decision	Strict policy	Address formally	Mentions CV
Type 1	0.937	0.222	0.172	0.000
Type 2	0.938	0.081	0.097	0.000
Type 3	0.936	0.105	0.105	0.000
Type 4	0.912	0.137	0.139	0.000
<i>p</i> -value	0.500	0.000	0.020	-1.000

Notes These numbers are calculated by assigning each individual to the pure type for which they have the highest estimated $\hat{\theta}_i$. Then, a conditional average is taken for each pure type. To test whether intentions differ across types, we compute the Mahalanobis (Wald) statistic and compare it to a χ^2_3 distribution. This relies on the multivariate normal approximation to the posterior described in Section 4.1 of [Gelman et al. \(2013\)](#). For non-binary outcomes, we first create a binary split at the median to split the sample.

N.2 What words describe responses?

We also ask the AI to describe each individual response using three adjectives. We replicate the analysis in section M.6 using these adjectives instead of tokens in the text response as inputs. Here, we report a word cloud describing the full corpus (Table A.28) before turning to the results of this exercise. First, Figure A.29 and A.30 report the results from this exercise pooling data from all vignettes providing respectively only positive, and positive as well as negative predictors. Figures A.31 to A.35 then report the positive predictors using only the adjectives for each individual vignette.

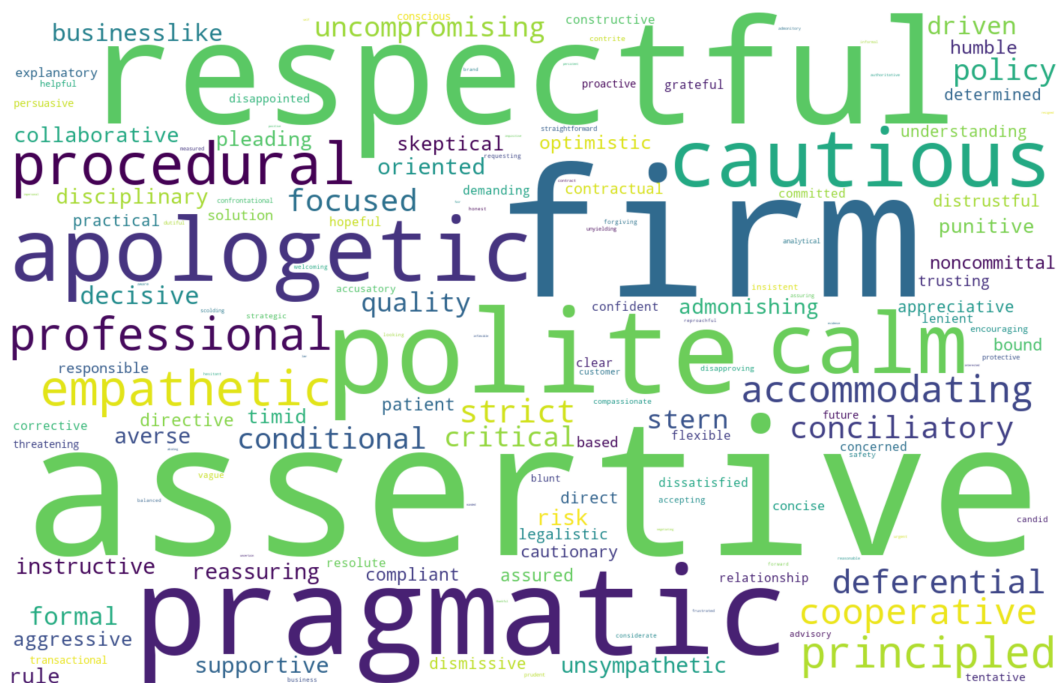


Figure A.28: Word cloud of adjectives

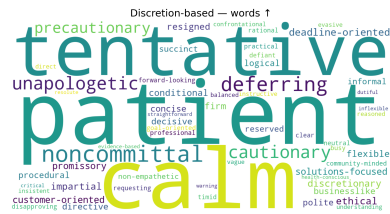
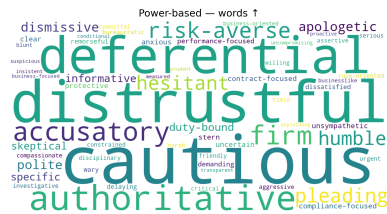
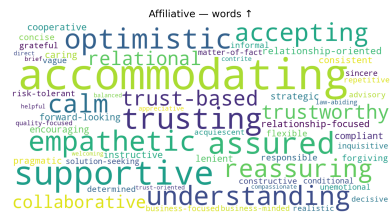


Figure A.29: All vignettes: Word cloud of adjectives that positively predict

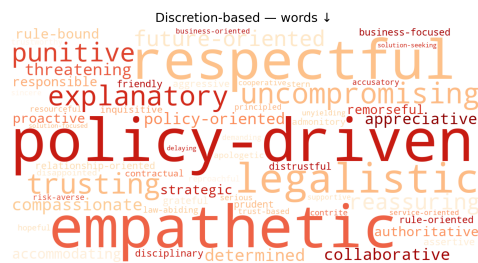
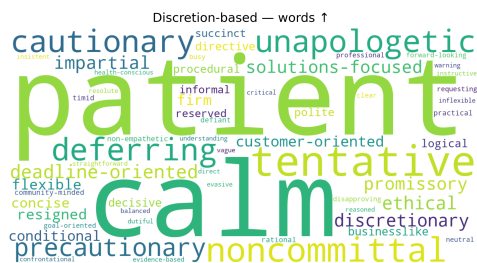
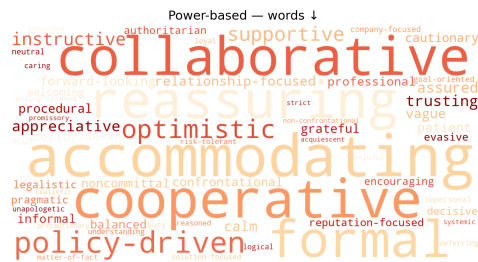
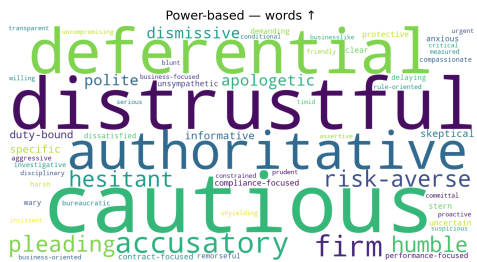
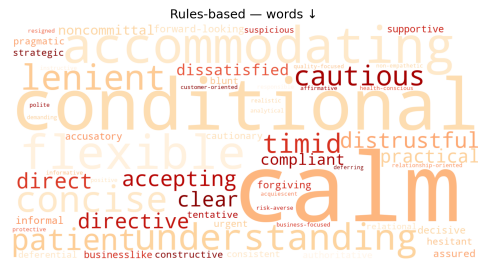


Figure A.30: All vignettes: Word cloud of adjectives that positively and negatively predict

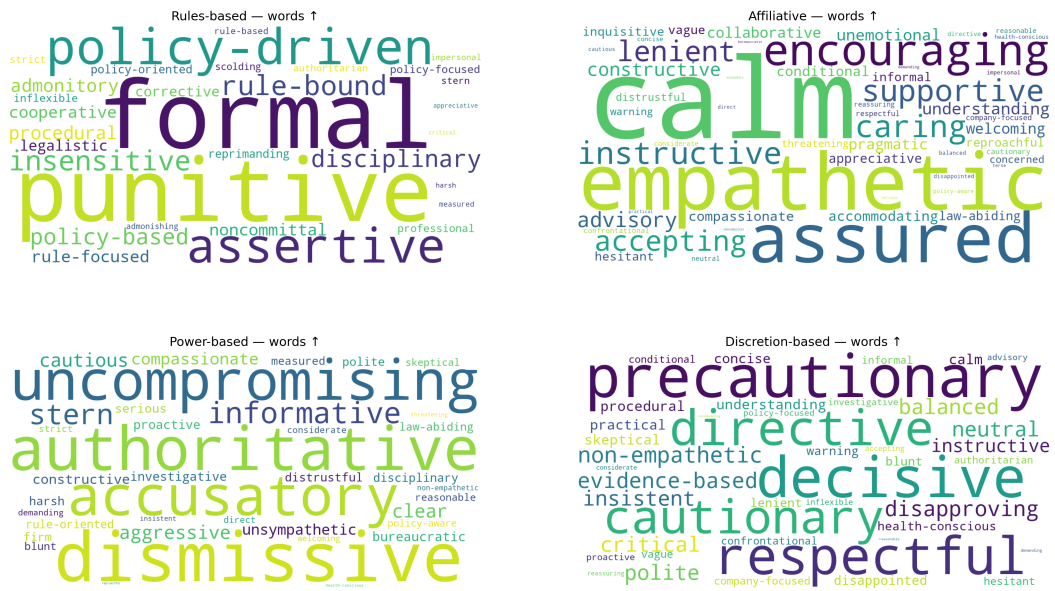


Figure A.31: Vignette 1: Word cloud of adjectives that positively predict



Figure A.32: Vignette 2: Word cloud of adjectives that positively predict

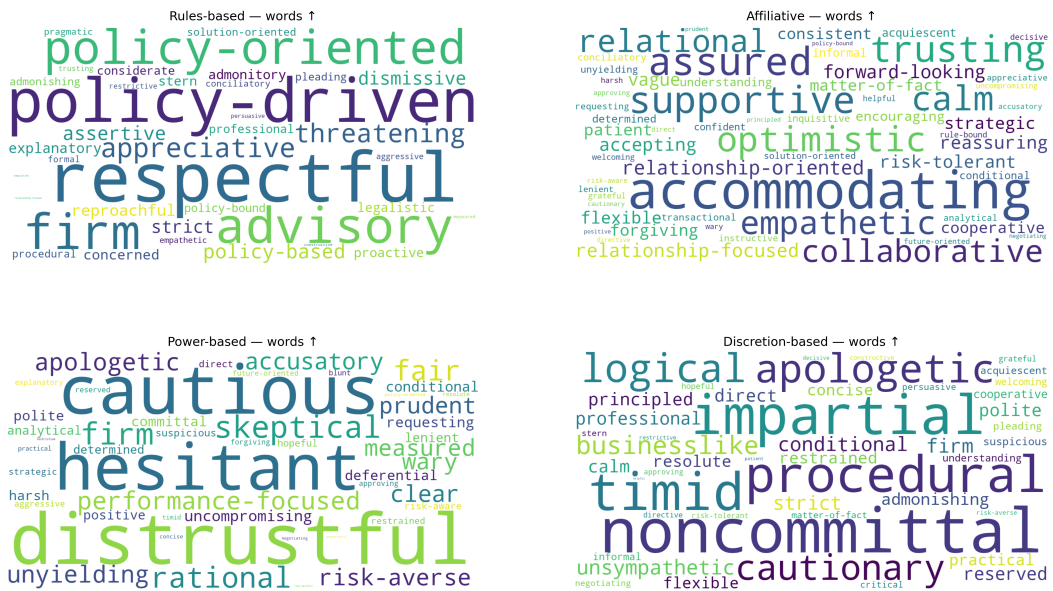


Figure A.33: Vignette 3: Word cloud of adjectives that positively predict

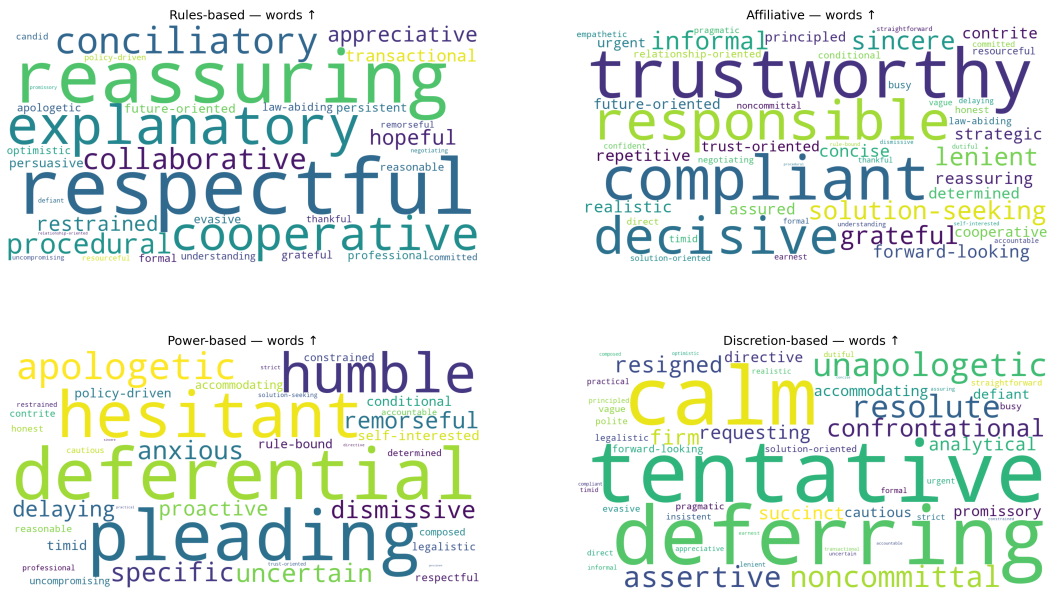


Figure A.34: Vignette 4: Word cloud of adjectives that positively predict

References

- Abebe, G., M. Fafchamps, M. Koelle, and S. Quinn (2026). Matching, management and employment outcomes: A field experiment with firm internships. *CEPR Discussion Paper DP21194*.
- Battaglia, L., T. Christensen, S. Hansen, and S. Sacher (2024). Inference for regression with variables generated by AI or machine learning.
- Bertrand, M. and A. Schoar (2003). Managing with style: The effect of managers on firm policies. *The Quarterly Journal of Economics* 118(4), 1169–1208.
- Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017). Stan: A probabilistic programming language. *Journal of statistical software* 76, 1–32.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* 1(3), 515 – 534.
- Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin (2013). *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.
- Gelman, A., D. Lee, and J. Guo (2015). Stan: A probabilistic programming language for bayesian inference and optimization. *Journal of Educational and Behavioral Statistics* 40(5), 530–543.
- Gelman, A. and D. B. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4), 457–472.
- Li, S. (2017). Obviously strategy-proof mechanisms. *The American Economic Review* 107(11), 3257–3287.
- Neal, R. M. and R. M. Neal (1996). Monte carlo implementation. *Bayesian learning for neural networks*, 55–98.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20(1), 53–65.
- Satopää, V., J. Albrecht, D. Irwin, and B. Raghavan (2011). Finding a “kneedle” in a haystack: Detecting knee points in system behavior. In *2011 31st International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pp. 166–171. IEEE.
- Steyvers, M., P. Smyth, M. Rosen-Zvi, and T. Griffiths (2004). Probabilistic author-topic models for information discovery. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 306–315.