The Importance of Being Earnest: Incentives, Noise and Hierarchy in Public Sector Decision-Making

ABSTRACT

Even high-performing organisations make mistakes, sometimes with major consequences, but much remains unclear about why errors arise, and how different kinds of error vary within organisations. I implement a simple survey experiment in highly selective and well-motivated public bureaucracy responsible for setting international development policy and programming. The experiment investigates the relative importance of random, non-systematic error (“noise”) and perverse incentives in generating organisationally sub-optimal decisions, exploring heterogeneity according to seniority within the organisation. I find that both noise and bad incentives have large effects. Noise is substantial, with error rates in a simple analytical task of 25%; bad incentives induce sub-optimal choices in an additional 16% of decisions. But while noise is a feature throughout the organisation, irrespective of seniority, the effect of bad incentives is ameliorated up the hierarchical chain: senior decision makers are unaffected. The importance of both noise and perverse incentives in understanding decision-making quality in a highly selective and motivated setting suggests limits to established approaches for improving organisational performance, and has implications for how decision-making structures in the public sector should be set.

Keywords:
Decision Making/Distributed Decision Making; Leaderships; Experimental/Laboratory Study
Introduction

The decision-making and advisory capacity of civil servants has enormous welfare implications. The middle and senior ranks of the civil service are consulted on virtually every matter of national importance in the UK, from foreign policy to taxation structure. In some departments civil servants may take discretionary spending decisions of up to GBP5 million (DFID 2016). I show that this decision-making is substantially affected by multiple sources of error in even the best-resourced bureaucracies. I argue that the standard economic prescriptions of better contracts and better selection are likely to have limited traction at the relevant margin. Instead, I show that organisational process choices, in this case around hierarchical decision-making, can substantially address these errors, if not eliminate them.

There is a long history of empirical and theoretical research into bureaucratic performance, dating at least to the 1950s in the work of Herbert Simon (Simon 1997; Simon, Smithburg, and Thompson 1991). Simon took a largely benign view of public bureaucracies, arguing that their function was limited by the cognitive capacity of their constituent members, and extended (or indeed sometimes further limited) by the organisational choices made in structuring these bureaucracies. Other scholars were more jaundiced. Much public choice theory is concerned with understanding the competing private and public incentives of bureaucrats, investigating how private returns, ranging from salary and promotion opportunities to opportunities for predatory rent-seeking, influence the decisions made by public agents, both political and bureaucratic (Besley 2007; Buchanan and Tullock 1962).

Empirical work has confirmed aspects of both of these views. Though little research yet exists in a public bureaucratic setting, it is becoming apparent that all organisations suffer from
considerable ‘noise’ (defined here as random error without a systematic bias in either direction) in decision-making: there is high variance in how decision-makers choose over the same choices, and considerable idiosyncratic error (Kahneman et al. 2016). This is very much in line with the spirit of Simon’s work on bounded rationality. It is also clear that cognitive performance (assessed in terms of organisational objectives) among public servants is also affected by their personal incentives to optimise their career prospects (Leaver 2009; Rogger and Somani 2018), though with considerable heterogeneity across ‘types’ of agents. Further research has demonstrated an additional cognitive limitation, one not predicted by the early work in organisational behaviour: like most other populations, civil servants fall prey to sunk cost bias, confirmation bias and other decision-making biases common in the behavioural economics literature (Banuri, Dercon, and Gauri 2017).

Related research has considered how best to motivate better bureaucratic and organisational performance. Theoretical foundations for optimally selecting and incentivising bureaucrats and public-spirited agents have been established (Acemoglu 2010; Besley and Ghatak 2005; Dewatripont, Jewitt, and Tirole 1999). Empirical testing has yielded mixed results, but the general point that there are returns to better selection and incentivisation of bureaucrats is clear (Ashraf, Bandiera, and Jack 2014; Bandiera and Lee 2015; Banuri and Keefer 2015). A related seam of empirical work has demonstrated the importance of management structures on performance (Rasul and Rogger 2013; Rasul, Rogger, and Williams 2017). However, virtually all of this work considers some form of project completion as the measure of bureaucratic effectiveness, rather than the quality of decision they make. This means we still have relatively little understanding of how the choices around contracting, selection and management affect complex decision-making and limited insight into the quality of policy making as opposed to
delivery. In this paper, I consider how hierarchical structures can affect organisational decision-making when contracting and selection are imperfect. Much of the literature on organisational hierarchy focuses on the contribution of CEOs and managers to firm value, their own behavioural biases, or the processes they follow and not their contribution to organisational cognition (Bloom and Van Reenen 2007; Lieberson and O’Connor 1972; Malmendier and Tate 2015). I investigate how hierarchy affects decision making both when decisions are difficult (involve interpreting non-obvious evidence) and when they are compromised (where bad incentives create a conflict between self-interest and probity). Additionally, I focus on middle- and senior managers, rather than CEO equivalents.

My contribution to this literature is threefold. Firstly, it is one of very few studies that experimentally examines decision-making in a well-functioning, highly selective and well-motivated public sector bureaucracy with major policy and spending importance, or among middle and senior managers. Secondly, the design of the study allows me to compare the effects of noisy decision-making and decision-making when incentives are misaligned with the organisational objective. And thirdly, I investigate how each of these effects varies with seniority in the organisation, which has important implications for how decision-making is best structured in such contexts.

I use a survey experiment to present an ideal set of evidence to civil servants, and give them a simple A/B choice over which to make an investment decision. For a treatment sample, I complicate this choice by lightly invoking a career incentive in the form of an expressed preference made by the most senior political figure associated with the Department – though this preference should have no bearing on the advice provided by civil servants. I find both noise and
bad incentives have substantial negative effects on decision-making – that is, idiosyncratic error is substantial, and made worse when an incentive to make a socially sub-optimal decision is introduced. However, both of these negative influences decline as we move up the hierarchy of the organisation. This effect appears to be at least partly driven by higher accuracy in the cognitive task assessed and willingness to challenge the Minister.

The paper proceeds as follows. The next section describes the setting. Section 3 describes the survey experiment. Section 4 provides the empirical strategy. Section 5 discusses sampling and attrition. Section 6 presents the results of the experiment. Section 7 provides some analysis of the mechanisms underlying the main result. Section 8 concludes with discussion considering the implications of these findings for how decision-making structures should be set in bureaucracies, and the limits to existing approaches for error minimisation.

Setting

The setting of my experiment is a high functioning Government bureaucracy with a strong international reputation. Empirical research into decision-making in such settings is still rare.

The Department for International Development is the UK Government Department charged with managing the UK’s GBP14 billion aid budget. It employs 2,754 staff (2,678.4 full time equivalents).\(^1\) The Department has a strong reputation internationally and plays an important role in international development, both directly through its bilateral footprint and its substantial contributions to multilateral bodies. DFID officials are influential in developing countries and in

the multilateral agencies which it funds. Officials also work across the UK Government, supporting national security and foreign policy objectives.

In common with the rest of the UK Civil Service, entry into DFID is highly selective. Entry into the civil service fast stream is an onerous process. It involves two online questionnaires, an e-tray exercise testing decision-making skills, a video interview before additional tests, sifting and finally a half-day assessment centre consisting of a leadership exercise, a group exercise and a written-analysis exercise before a possible final assessment. In addition, DFID requires that all technical specialists hired at any level have at least a Master’s degree in a relevant discipline. Those joining the Department at a more senior level are usually expected to have eight or more years of experience, a Master’s degree in a subject relevant to the role, substantial technical expertise in the area relating to the role and a strong history of work experience. A lengthy and complicated application form is followed by at least one panel interview, usually with three interviewers, one of whom may be from another Government department to ensure the person selected has the minimum qualities required to work in any Government department, not just the role for which they are interviewed. All staff are required to demonstrate basic numeracy through testing. The result is that DFID’s HR structure provides rigorous competition and selection and protection of technical standards.

Furthermore, all staff are covered by the Civil Service Code, which provides both formal protections and aspirational statements of conduct to civil servants. The code is protected by the

---

2 www.faststream.gov.uk/faqs/
Civil Service Commission. These aim to protect their ability to provide impartial advice to Ministers and make decisions in the public interest. The civil service code states:

“As a civil servant, you are appointed on merit on the basis of fair and open competition and are expected to carry out your role with dedication and a commitment to the Civil Service and its core values: integrity, honesty, objectivity and impartiality. In this code:

- ‘integrity’ is putting the obligations of public service above your own personal interests
- ‘honesty’ is being truthful and open
- ‘objectivity’ is basing your advice and decisions on rigorous analysis of the evidence
- ‘impartiality’ is acting solely according to the merits of the case and serving equally well governments of different political persuasions

These core values support good government and ensure the achievement of the highest possible standards in all that the Civil Service does. This in turn helps the Civil Service to gain and retain the respect of ministers, Parliament, the public and its customers.”

The Department has a hierarchical structure, with work produced in teams typically made up of staff of the same grade or closely bunched grades and submitted and ‘cleared’ up the hierarchical chain until a final decision is taken, though the level of seniority at which a policy or spending decision can be approved varies by importance of the decision. Throughout this paper, I define

---

3 “If a civil servant is asked to do something which conflicts with the values in the Code... Their department should investigate their concern. If the Civil Servant remains dissatisfied following the outcome of the investigation, they may bring a complaint to the Civil Service Commission. In some cases, the Commission may also hear a complaint direct.”

hierarchy as grade seniority in the civil service structure. While some people do hold high grades without line management or spending responsibility, there is a clear correlation between decision-making responsibility and grade.

Staff report a high degree of intrinsic motivation, but dissatisfaction with pay. After many years of below-inflation pay rises, achieving a real-terms pay rise is typically only possible through promotion or a posting overseas.

The experiment
I designed a survey experiment that tests both the application of basic cognitive skills and how these hold up against compromised incentives. The aim of this experiment was to better understand how well civil servants process evidence; if signals from powerful political leaders disrupt this; and how the hierarchical decision-making structure of the civil service amplifies or mitigates these effects. It is useful to distinguish at the outset between underlying cognitive ability and the error rate in specific cognitive tasks. This experiment investigates the latter, not the former. Even though the setting features highly selective recruitment, we would still expect error rates to be non-negligible as this has been observed in other highly motivated settings, from both idiosyncratic error (Kahneman et al. 2016) and systematic bias (Thaler 1987). Errors of any kind matter in civil service settings, since the advice of civil servants is typically used as a benchmark against which to hold elected Ministers to account; and Ministers correspondingly

5 Over the last ten years, the annual Civil Service People Survey's show that satisfaction and identification with the Department’s mission and objectives is usually between 86-92% while satisfaction with pay and conditions hovers around 36-42%. See: http://www.gov.uk/government/publications/department-for-international-development-civil-service-people-survey-2018
place great weight on the advice provided by their civil servants. The hierarchical structure of decision-making ensures that decisions are usually ‘cleared’ by more senior staff after the initial analysis is undertaken by more junior civil servants.

The experiment took the form of a vignette, asking the respondent to imagine that they were assessing options for a new DFID programme in the field of education. They are explicitly informed that the programme’s objective is to improve learning outcomes. They are told that they have commissioned research in the place in which the programme will be run, performing an A/B comparison of two kinds of intervention by means of a randomised control trial (RCT), and given the results of this RCT. They are then asked to simply choose which of the two interventions they would select, noting that they will personally sign the submission of the business case to the Secretary of State. In the control group, this is the extent of the problem. The numerical skills required to correctly select between the two options are well below the threshold of a generalist civil servant. Further details are provided in Appendix A.

The treatments were identical except for the addition of a new paragraph, providing information on Ministerial preferences while reiterating that improved learning outcomes are the objective:

“The Secretary of State has been giving a series of interviews in which she has stressed the importance of curriculum development [OR teacher incentives] for improving learning outcomes, which she has said is one of her key objectives.”

Using this additional text, I created two treatment groups. Treatment group 1 were given a paragraph of information in which the Minister has publicly proposed the less effective
intervention. Treatment group 2 receive a paragraph in which the Minister backed the more effective intervention (which intervention was more effective was also randomised).

The vignette reflects the kind of choice often made in ‘spending’ departments, where decisions have to be made to between alternative programmes or investments, or to select the most cost-effective way of achieving some specified output. In DFID, almost all spending decisions are made on the basis of some form of cost benefit analysis comparing different approaches to achieve pre-specified objectives. As such, the vignette captures a salient form of decision-making, and learning more about how mistakes can be made in this kind of decision-making is valuable for improving Government spending decisions.

While hierarchy was not explicitly invoked in the vignette, we collected information on the seniority of the respondent to allow investigation of how this affects outcomes. One possibility is that those at more senior levels are more likely to be influenced by a desire to please Ministers because they have more regular interaction with them, or are more likely to be in the running for the few highly senior jobs that involve some Ministerial say in appointments. Another possibility is that senior staff are less likely to act to please Ministers because they more closely adhere to the civil service code (which may explain their original promotion) or because they are more secure in their jobs, having already secured influence, pay and status. Interacting treatment with seniority will provide insight into this.

Inference of the results depends on two assumptions: firstly, that despite the unincentivised nature of the experiment, respondents make a serious attempt to solve the computational problem
in the vignette. And secondly, respondents recognise the significance of personally signing the submission and the size of the business case.

The first is necessary to use the accuracy of choice in the control group to make inferences about the rate of idiosyncratic error in the organisation (that is, we need to assume that the results of the survey are a reflection of the actual propensity to make errors of the respondents). The second is required if we are to infer an effect of personal incentives to please the Minister on decision-making from any difference between treatment and control groups. This requires that respondents realise that the decision taken in the vignette would not be anonymous but personally attributable. Two pieces of evidence suggest that these are reasonable assumptions. Firstly, the observed rates of correct choices are in line with those found in similar populations backed up by questions testing attention (Banuri et al. 2017). Secondly, the survey was piloted with a group of out-of-sample staff, all of whom recognised the personal responsibility signalled in the vignette.

A more general assumption is that we can learn something about the behaviour of bureaucrats in their daily work from a survey experiment. It is indisputable that a survey is a fundamentally different context to normal working processes. However, surveys are not infrequent in the civil service, and it was complete at work and the vignette was designed to mimic regular choices made in the normal course of work. The vignette is likely to be less pressured than regular work, and as such may approximate an ideal work environment.

---

6 This reflects actual practice. Submissions are signed and responsibility for their contents lies with the sender.
7 We asked how they went about comparing the two interventions, and what, if any, importance they attached to the fact that they would be author of the covering submission for the business case.
Empirical strategy

I use the Linear Probability Model (LPM). Its major advantage is in the straightforward interpretation of the coefficients of interest, which can simply be read as percentage point changes in the likelihood of the event of interest. In the case of this experiment, coefficients on variables of interest represent the increase (or decrease) in the likelihood of selecting the more effective option from a one-unit increase in the variable in question.

The basic models estimated are as follows:

\[
(1) \quad Y_i = \beta_0 + \beta_1 Treatment1 + \beta_2 Treatment2 + \epsilon \\
(2) \quad Y_i = \beta_0 + \beta_1 Treatment1 + \beta_2 Treatment2 + \beta'X' + \epsilon 
\]

In this formulation, \( Y_i \) is the variable of interest, the probability of selecting the more effective option. \( \beta_0 \) is an intercept term, interpreted as the baseline probability of selecting the correct option (i.e. the probability holding all control variables at 0). \( \beta_1 \) and \( \beta_2 \) are the effects of the two experimental treatments (the Minister supporting the more or less effective option, respectively), with the coefficients giving the percentage increase or decrease in the likelihood of selecting the right option under that treatment. \( \beta'X' \) is a matrix of control variables used to improve accuracy of the model. \( \epsilon \) is the error term.

To investigate how these effects vary with position in the civil service structure, a third specification will also be estimated, to investigate the interaction between treatment and an important organisational characteristic: hierarchy. Hierarchy is defined here as seniority in the civil service grading structure.
\( Y_i = \beta_0 + \beta_1 \text{Treatment1} + \beta_2 \text{Treatment2} + \beta_3 \text{Treatment1} \times \text{Seniority} + \beta_4 \text{Treatment2} \times \text{Seniority} + \beta' \mathbf{X} + \epsilon \)

This is my preferred specification. Empirically, ‘seniority’ will be defined as a simple 0/1 dummy variable capturing whether a staff member is part of the middle and senior management grades of the civil service or not. In this specification, the X variables will also include seniority. The interaction terms will give the impact of the two treatment arms among senior as opposed to junior staff. This will allow us to assess whether the causal effect of invoking incentives is different at different grades/levels of seniority.

If we accept the assumptions underlying the experiment, we can infer the:

1. \( \beta_0 \) in specification 1 provides an indication of the frequency of errors in one application of cognitive ability, with no incentives invoked and without controlling for individual characteristics. This is essentially the mean of the dependent variable in the control group.

2. \( \beta_1 \) gives the causal impact on decision-making quality of information that the Secretary of State has backed the less effective option. Any effect could reflect a belief that promotion prospects are harmed by contradicting the Secretary of State or simple disutility from disagreeing with Ministers, either through stress or a desire to please.

3. \( \beta_2 \) gives the causal impact of information that the Secretary of State backs the more effective option, indicating if agreeing with Ministers has a career or other utility benefit.

4. \( \beta_3 \) and \( \beta_4 \) give the causal impact of seniority on the two treatment effects above: does the information have a different impact on junior or senior staff?

---

8 In annex B, various alternative coding choices are discussed. All preserve the core results.

9 Note that this need not operate via political interference, since senior civil servants may penalise ‘difficult staff’ even if Ministers make no such representations to them.
Sampling and attrition

The population the survey sample was drawn from was all staff likely to make some programming or policy decisions and use evidence in their work. The population was then stratified by position in the hierarchy and professional specialism. The participants were randomly selected from each stratum into the survey sample. Assignment of the survey sample into control and treatment arms was random and not stratified. All were sent the survey link to complete within a 6-week deadline.\textsuperscript{10}

Our response rate was 61% (604 individuals). At this stage attritors will not have seen any part of the survey and will be unaware of the existence of the experimental vignette, let alone to which treatment group they are assigned. Nevertheless it is possible that certain kinds of civil servant were systematically less likely to respond to the survey, which would compromise external validity – the extent to which the results presented here can be generalised to the entire population of civil servants in DFID.

To see if this was the case, I ran a regression to see if any respondent characteristics predicted their likelihood of opening the survey at all. No respondent characteristics predict attrition except tenure and whether or not they were senior manager. For tenure the coefficient is small – equivalent to a 0.4% increase in likelihood of opening the survey per year of service. With a mean tenure of around 7 years in the sample, this suggests a slight overrepresentation of more experienced staff in the results, but the size of this effect means it is of little practical

\textsuperscript{10} Details in Annex A.
significance. Senior managers were more likely to respond to the survey, which does suggest that they are over-represented in the sample compared to junior staff, but I control for grade in the main specifications, which should minimise the extent to which this compromises generalisation to the whole population. Full results are presented in Appendix B.

Among survey participants, a small number opened the survey but did not complete it. Of the 604 who opened the survey 536 completed it (89%). Of those who did not complete the survey, none progressed as far as observing the experimental vignette\(^{11}\). A regression of attrition on treatment arm yields no significant relationships, indicating that attrition is not predicted by treatment arm.\(^{12}\) Chi-square tests confirm no differential attrition by treatment arm either at either response or completion level (i.e. there is neither differential attrition at the stage of opening the survey at all nor among those who completed the survey and the experimental vignette). The 989 sampled and those responding were divided into the control and treatment groups as follows:

<table>
<thead>
<tr>
<th></th>
<th>N (assignment)</th>
<th>N (respond)</th>
<th>N (completed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>330</td>
<td>199</td>
<td>173</td>
</tr>
<tr>
<td>Treatment 1: Minister backs the less effective option</td>
<td>327</td>
<td>203</td>
<td>188</td>
</tr>
<tr>
<td>Treatment 2: Minister backs the more effective option</td>
<td>332</td>
<td>202</td>
<td>175</td>
</tr>
</tbody>
</table>

Calculating statistical power \textit{a priori} was guesswork to some extent, given that there was no publicly available data from which to judge the baseline ability to interpret evidence, effect size for treatment or standard deviation in interpreting evidence for the population. However,

\(^{11}\) Since certain pages required mandatory information to be completed before the respondent could move on to the next page, it’s possible to assess whether or not it was possible for the respondent to have observed the vignette by looking at how much data they did enter. Almost all failed to enter any data at all, and those that started the survey all exited it before reaching the main body of the survey, let alone the vignette.

\(^{12}\) This is true whether a logistic or LPM model is used to estimate the regression.
assuming around 75% of the population select the correct rate in the control group, the study was powered to detect relatively small effect sizes – of 5 percentage points - 80% of the time if the standard deviation was as high as 17.

Results

Descriptive data from the survey suggest that the idea that civil servants would rather avoid challenging their Ministers directly is sound. Though the data doesn’t directly measure the disutility of challenging Ministers, but we can compare how ‘comfortable’ respondents report being when challenging Ministers compared to civil servants.

Figures 1 and 2: How comfortable are you challenging the use of evidence by a…

Clearly, there is a substantial difference here: almost all respondents are ‘somewhat’ or ‘very’ comfortable challenging civil servants on poor use of evidence, but almost as many report being ‘not at all’ as ‘somewhat’ comfortable challenging Ministers. This is not, since civil servants may still do what they feel uncomfortable doing, but it does suggest that challenging a Minister carries a personal disutility, while the data are also consistent with perceptions of a career penalty to challenging Ministers. Empirical studies do suggest that career incentives do affect performance
(Bandiera and Lee 2015; Dewatripont et al. 1999; Leaver 2009). This experiment will test if the relative lack of comfort civil servants feel challenging Ministers (whether driven by career concerns or by simple disutility) also affects the advice they provide to them.

Table 2: Effect of information on Ministers’ preferences on selection of the most effective investment option

<table>
<thead>
<tr>
<th>Dependent variable: Selected the more effective option</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment 1: Minister backs less effective treatment</td>
<td>.0105768 (.0463862)</td>
<td>.002119 (.0451754)</td>
<td>-.1580993* (.0946419)</td>
<td>-.1494653 (.0957531)</td>
</tr>
<tr>
<td>Treatment 2: Minister backs more effective treatment</td>
<td>.025896 (.0467204)</td>
<td>.0072463 (.046237)</td>
<td>-.1106508 (.1027822)</td>
<td>-.1066987 (.1024238)</td>
</tr>
<tr>
<td>Treatment 1 x Seniority</td>
<td></td>
<td></td>
<td>.2222653** (.106911)</td>
<td>.2139339** (.1083265)</td>
</tr>
<tr>
<td>Treatment 2 x Seniority</td>
<td></td>
<td></td>
<td>.1605368 (.1147056)</td>
<td>.1596152 (.1146267)</td>
</tr>
<tr>
<td>Constant</td>
<td>.734104*** (.0336845)</td>
<td>.4817494*** (.0778528)</td>
<td>.5747385*** (.0914616)</td>
<td>.5309702*** (.1014838)</td>
</tr>
<tr>
<td>Controls</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Comfort challenging Ministers controlled?</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0006</td>
<td>0.1103</td>
<td>0.1193</td>
<td>0.1278</td>
</tr>
<tr>
<td>N</td>
<td>536</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses * Statistically significant at the 10% level ** Statistically significant at the 5% level *** Statistically significant at the 1% level

Basic controls: a dummy variable taking the value 1 if the respondent is a middle or senior manager and 0 if not; a dummy variable coding whether or not the staff member was recruited in country as opposed to being put through the full recruitment process in the UK; a dummy variable for whether or not the respondent self-assesses as having high ability in interpreting quantitative evidence (levels 4 or 5 from the 5 point likert scale); a dummy variable for whether teacher incentives was the ‘correct’ choice; and gender, tenure and professional cadre.

Table 2 presents the results of the basic LPM regression without controls (specification 1), with additional basic controls (specification 2), with basic controls and interactions with seniority (specification 3), and with basic controls, interactions and additionally controlling for self-
reported comfort challenging Ministers (specification 4). The next two sub-sections discuss what these results tell us about noise and bad incentives as sources of error, respectively.

*Noise*

The constant in specification 1 gives the percentage of respondents who selected the more effective option in the control group, when they had no information on Ministerial preferences and no incentives to select anything other than the better option.\(^\text{13}\) This is exactly equivalent to the mean of the dependent variable in the control group. This rate is just 73 percent, meaning that around one quarter of respondents were unable to interpret the two simple sets of figures provided, made a mistake in doing so or were unwilling to update any prior held in favour of one or the other option. This is despite no common cognitive bias being primed and the explanatory information explicitly informing them that the evidence was collected at their own behest, in their precise context in which they will be running their programme and that the data are directly comparable. In specification 2, I control for seniority in the organisation, as well as a range of other characteristics. Seniority does not predict likelihood of selecting the better option, suggesting that noise is a feature throughout the hierarchy of the organisation.

Since the vignette approximates an ideal set of information, presented clearly and simply, this error rate suggests that idiosyncratic error is common even in populations with good education, careful selection, good remuneration and substantial experience (a mean of 7 years in the job).

\(^{13}\) In each specification, the constant is interpreted as the baseline rate at which surveyed staff selected the more effective of the two learning interventions, with all other variables held at 0. Only the coefficients on the two treatment arms, and interactions with them, can be interpreted as causal since assignation to treatment was random. Other coefficients may have explanatory value, but we cannot be sure that any relationship is causal.
Just this simple result lends support to the nascent literature on noise in organisational decision-making. Recent research has attempted to quantify the size of this effect and suggested mitigation through the use of algorithms and automation (Anderson, Kleinberg, and Mullainathan 2016; Kleinberg et al. 2018), simple decision guides or checklists (Gawande 2010), or new decision-making techniques (Kahneman, Lovallo, and Sibony 2019). I discuss alternative and complementary approaches in the discussion.

**Bad incentives**

Neither specification 1 or 2 find any effect from either treatment on the rate at which officials select the more effective of the two interventions. Once we introduce heterogeneity by seniority/grade, however (specification 3), large effects are found. This is only true of treatment 1 – when the Minister signals support of the less effective option. In other words, while (less senior) civil servants may suppress their reading of the evidence for fear of contradicting a Minister, there is no corresponding boost from supporting a Minister.

The effect of treatment 1 is substantial. At the lowest grades in our sample, respondents are 16 percentage points less likely to select the more effective option when the Minister has publicly spoken in support of an alternative approach. This effect is both statistically significant at the 10% level (p=0.095) and practically significant – it is the equivalent of introducing additional errors to more than a sixth of decisions.

However, this effect is mitigated by the interaction of the treatment with position in the hierarchy. Specifically, the regression shows that middle and senior managers are 22% more likely to select the more effective option than their junior colleagues under treatment 1 (p=0.038). It appears that
senior civil servants are less likely to be influenced by political preferences in the face of evidence than their more junior colleagues. Further, we cannot reject that $\beta_1 + \beta_3$ jointly sum to zero ($p=0.2015$). In other words, for senior officials, knowledge that the Minister has backed a less effective option does not affect their decision-making quality either way, neither inducing them to make more mistakes nor inspiring additional attention and hence greater accuracy. This is a highly encouraging finding, suggesting that the decision-makers most close to Ministers are those most likely to challenge them when necessary, but are not unduly confrontational.

This encouraging finding does not appear to be an artefact of the experiment, driven by respondents guessing the intention of the survey and adjusting their answers accordingly. Respondents might ‘game’ the experiment if they guess that, at mention of the Minister, the vignette is testing willingness to challenge them. If so, they may contradict the Minister in their response regardless of what they would do in real life. The results outlined above might be the result of senior staff gaming the experiment more effectively. One way of testing whether or not this is the case is looking for signs of gaming by comparing the results of treatment 1 and treatment 2 among more senior staff, or among those most willing to challenge Ministers (as we will see later, willingness to challenge Ministers is closely correlated with seniority).
If gaming were prevalent among the higher grades, we would expect that those most comfortable challenging Ministers respond to information about the Minister’s preference by contradicting them. As such, they may select the opposite option to the Minister irrespective of treatment, leading to a higher rate of correct choices in Treatment 1 and a lower rate in treatment 2.\(^{14}\)

The raw data suggests some such effect is observed (Figs 3 and 4). However, it is very small, and a Fisher’s exact test finds no significant difference between the rate of selecting the more effective option among those very comfortable challenging Ministers across arms (Fishers Exact = 0.665). The observed difference in Figures 5 and 6 may therefore be a simple artefact of the data. Similar results are obtained when we look at seniority directly: there is a small dip in accuracy among the most senior civil servants in Treatment 2, but the effect is small and statistically insignificant. The key result, that needing to contradict the Minister lowers accuracy

\[^{14}\text{This also helps explain why treatment 2 has no significant impact: it appears to only matter for those who know they are not at all or only somewhat comfortable challenging Ministers.}\]
among junior staff, but has no effect on those most senior and likely to be in contact with them, does not therefore appear to be driven by gaming of the experiment.

**Mechanisms**

There are two immediately obvious reasons we might see a diminishing effect of treatment with seniority. The first is that the public sector is effective in promoting people with higher cognitive ability (at least in the interpretation of quantitative evidence, tested here), and this cognitive ability is especially important under pressure (as extra information about the Minister’s preference might constitute, even if it has no incentive value). The second is that there is a relationship between position in the hierarchy and ability to challenge Ministers. This relationship could be causal, with promotion providing the institutional mandate, job security or training to contradict Ministers when necessary. It might also be selective, with the civil service identifying the most independent-minded individuals to promote, so individuals who are inherently more willing to challenge are most likely to be promoted.

If the civil service is able to identify and promote the most cognitively able individuals, we should see that when no incentive complications are introduced, more senior respondents do better at selecting the most effective option.
The raw data do not bear this out. In Figure 5, there appears little difference in selection of the correct option in the control group by seniority, as indicated by the almost entirely overlapping confidence intervals. Indeed, a regression of selecting the better option on seniority and controls, restricted to the control arm, returns no significant coefficients (n=173). It is not, therefore, that more senior staff are systematically more able to select the more effective option. However, it may be that ability matters particularly when stress is greatest – i.e. when the Minister is involved. We can test this by adding an interaction between ability and treatment to specification 3. This provides evidence that the seniority effect is partly driven by ability under stress: while the coefficients and significance of treatment 1 barely change, the interaction effect between seniority and treatment falls slightly (from 0.22 to 0.18) and loses significance (from $p=0.038$ to $p=0.13$). However, the coefficient on the interaction between ability and treatment 1 is still insignificant. Ability under pressure is part of the explanation, but not the primary mechanism.
A second possible reason we observe treatment effect varying with position in the hierarchy is that promotion has a causal effect on the ability of civil servants to challenge Ministers, or is awarded to those most willing to do so. The survey provides some useful data here. All respondents were asked to report their self-assessed comfort in challenging Ministers; for this explanation to have weight we would expect a clear correlation with position in the hierarchy.

**Figure 6: Average comfort challenging Ministers by civil service grade (0=don’t know, 4=very comfortable)**

Figure 6 demonstrates that this is exactly what we observe: a clear step-wise increase in willingness to challenge Ministers as we move up the civil service grades (confirmed by regression analysis with a full set of controls). We can also test whether this comfort in challenging Ministers is driving the results observed in specification 3 by including it as a control variable in the regression. It is omitted in the preferred specifications precisely because it was expected to be at least one of the mechanisms through which effects would be observed (i.e. respondents select sub-optimal options because they are unwilling to challenge the Minister).
Specification 4 includes this control. There is only a small effect from controlling for self-reported comfort confronting Ministers. The coefficient on treatment 1 over the sample falls slightly, corresponding to a reduction in the effect of treatment of 1 percentage point or so; and the coefficient on the interaction with hierarchy also falls by around 0.8 percentage points. These effect sizes are small, though in the direction expected. One explanation for these relatively small effects might be that self-reported comfort is a weak measure of actual willingness to challenge Ministers. Alternatively, it may be that comfort challenging Ministers only matters when Ministerial opinions are salient to the decision. However, adding an interaction term between a dummy variable capturing high comfort challenging Ministers with treatment is insignificant and has almost no impact on the coefficients of interest in specification 3, nor their significance.

If ability and comfort challenging Ministers aren’t the primary drivers of better performance of senior managers in this task, what is? Though untestable from the data available, it is tempting to return to the civil service code, discussed earlier. The code is most familiar and salient to the more senior civil servants who most routinely find themselves giving advice to Ministers, either in person or in writing. Figures 1 and 2 clearly demonstrate that there is something more difficult about challenging Ministers. It is striking therefore, that among the senior decision makers surveyed it is not just those with high ability or least uncomfortable with challenging Ministers who take this step, but most of them. The idea that an ethos that privileges impartiality and ‘doing the right thing’ even at a personal cost may trump self-interest or discomfort isn’t new: Oliver Williamson suggested that probity would be more important than performance incentives in ‘sovereign transactions’ such as foreign policy (Williamson 1999). More recently, economists have increasingly investigated morality, identity and norms as determinants of economic behaviour and choices (Akerlof and Kranton 2010; Bowles 2016). Use of a code to encourage
pro-organisational behaviour is in line with these ideas. The next section discusses the implications of these findings.

Discussion

Organisational performance is a product of not only the selection and incentivisation of individuals, but the management processes they use and the ways in which they are combined. In some sense, the selection and incentivisation of staff determines (some of) the inputs of an organisational production function (especially talent and effort), while the management of them, the structures through which they interact and the ways in which they work together are the technology with which organisational outputs are produced.

While we know an increasing amount about how best to select staff and incentivise their efforts (Bandiera and Lee 2015; Banuri and Keefer 2015; Besley and Ghatak 2005). But in elite organisations, which already make substantial investments in selection and contracting, how likely is it that we can find meaningful margins on which to improve? In our setting, we find that even after selection processes that involve multiple interviews and tests, group work and an explicit numerical assessment, a public organisation with a strong international reputation still sees around a quarter of decision-makers select the less effective option when the only computation necessary is the calculation and comparison of two ratios. At the individual level, we can expect basic errors to be made even in the best-functioning kind of organisation.

It may be necessary to accept that there is no feasibly adopted contract or selection process that completely eliminates error. This approach underlies the recent research into the use of algorithms (Anderson et al. 2016) and process technologies for error minimisation (Gawande
2010; Kahneman et al. 2019; Sibony, Lovallo, and Powell 2017). These approaches do not eradicate error; they merely provide an additional layer of process to reduce it. At worst, some can magnify bias, though there are ways to guard against this (Mullainathan and Obermeyer 2017). Further, few of these approaches are capable of handling incentives to select sub-optimal decisions, which this paper argues often coexist with noise.

While these approaches tend to focus on the individual decision or decision maker (indeed some aim to eliminate them), moving to a genuinely organisational approach to addressing errors generates different kinds of solutions. These solutions are built in to organisational structure and complement rather substitute for the approaches Gawande, Kahneman, Mullainathan and others observers put forward. I relate some of these solutions to the results of my experiment now.

Firstly, a resilient organisational decision-making process may need to build in redundant capacity or cognitive function, an approach that goes against the efficiency-maximising ethos of management consultancy and mainstream economic thinking on industrial organisation. The errors observed even in the control group suggest that mistakes may be inevitable. While some of the error rate can be explained by confidence in quantitative skills or professional background, most cannot. Therefore, optimising the mix of skills and backgrounds with respect to the kind of decisions the organisation typically faces may contribute to reducing the rate of individual error, but it is unlikely to eliminate it. Using groups to make decisions may make an additional dent in the error rate however, either through discussion and explicit collective cognition or through voting or group selection of a desired option. Since, in our example, those making errors are in a minority, simple majority voting would result in an increase in decision-making quality even if
teams of three were constructed at random.\textsuperscript{15} A number of studies find improvements in cognition at the group level (Ahmed 2017; Charness and Sutter 2012; George and Chattopadhyay 2008; Kao and Couzin 2014). It may pay to use groups to make decisions even when an individual can process the information alone. This does, of course, raise other issues, not least the possibility of groupthink if groups are constructed from like-minded raw material (Janis 1982).

Another approach is to explicitly build redundancy into the decision-making process, overlapping individual remits in such a way that any cognitive process is performed by at least two individuals before contributing to a final decision. In Edwin Hutchins’ classic anthropological study of organisational decision-making, Cognition in the Wild, this is exactly the structure adopted by the navigational team he studied (Hutchins 1995). He examined how cognition proceeded both in routine circumstances and under conditions of high stress. His study demonstrated that particularly under pressure, ‘redundant’ cognitive capacity played an important role, with errors being picked up and questioned before a final decision was taken.

Secondly, decisions made by technically skilled and competent staff may still benefit from passing through a hierarchy before finalisation as a way of identifying and mitigating errors. Though this kind of ‘slow thinking’ is often held up as a problem of bureaucracy (when used as a pejorative), it may pay cognitive dividends through three channels. Firstly, it provides another way of creating ‘cognitive redundancy’ of the type Hutchins identified. Secondly, it can improve accuracy if the senior member of staff is more cognitively able under pressure; this may be more efficient if checking decisions allows the more sophisticated staff member to influence more

\textsuperscript{15} Assuming, of course, that the simple act of being in a group does not change their behavior, which some research finds evidence for (Charness and Sutter 2012).
decisions than doing them from scratch. And thirdly when personal incentives may run counter to
the best choice for the organisation and stakes are high, it can provide an organisational buffer to
the effects of misaligned incentives. 16 While research has examined the overall value of CEOs
(Lieberson and O’Connor 1972), their biases (Malmendier and Tate 2015), the role of hierarchy
in coordination or convening power (Anderson and Brown 2010) or in determining the
boundaries of the firm (Hart and Moore 2000), the role of hierarchy in organisational cognition
has received less attention, but may be a significant source of value. 17 This echoes early work in
which managerial skill is a multiplicative term in a firm’s production function (Lucas 1978;
Rosen 1982).

Cognitive redundancy is not an explicit objective in the structural choices of many organisations,
including the civil service, but it is an unintended consequence of hierarchy. Given the overlap in
the skills necessary for selection into a civil service job, there is likely to be cognitive redundancy
both horizontally (among colleagues) and vertically (between leaders and their staff). The
evidence provided by my experiment suggests that both are likely to contribute to better decision-
making on pure cognitive grounds, since a minority of staff fail the simple cognitive test set here,
and the failure rate gently declines with seniority. Even more importantly, this experiment
suggests that cognition under stress is also improved by cognitive redundancy, since some
portion of that redundancy is accounted for by overlapping cognitive function between actors
who have different attitudes to challenging Ministers.

16 Of course, it’s also possible that in some cases the malign incentives will affect only the senior
staff and not the more junior.
17 Though, as James March has observed, the exact identity of these managers may not matter,
simply that they are chosen from a pool of near-identically competent candidates. See:
Ultimately, all organisations are likely to be imperfect in multiple ways. From the beginnings of the theory of the firm (Coase 1937; Williamson 1985) and theories of public administration and organisational behaviour (Gibbons 2013; Simon 1997), emphasis has been placed on how organisations arise in response to imperfections in markets, information and contracting, and in turn may be characterised by multiple imperfections themselves. The true challenge is to recognise these various imperfections and to adapt organisational structures and processes to mitigate them. My experiment finds negative effects from introducing any information of Ministerial preferences on decision-making. Indeed, it even suggests that decision-making is compromised when the Minister is correct. While the first-best policy response would be to somehow limit Ministers to articulating impacts to be achieved rather than the means they prefer, limiting political expression is next to impossible (and undesirable on democratic grounds). Rather, the civil service promotes the most cognitively able individuals and encourages a culture of independence and challenge through the civil service doe to try and limit the effects of Ministerial preferences on how civil servants provide advice. This is an organisational adaptation to an insurmountable organisational imperfection. The civil service code, and its supporting infrastructure may, then, be playing an important role in cognition.

---

18 The point estimate on accuracy of civil servant decision-making is negative even when the Minister is correct, though not statistically significant.
Annex A: The Experiment

I took advantage of the regular but infrequent ‘DFID Evidence Survey’, a wide-ranging survey of a representative sample of DFID staff. The survey is undertaken every two years and took around thirty minutes to complete. It collected basic information about roles, seniority, professional and technical qualifications and self-assessments about their level of ability in interpreting evidence.

It then moves on to a series of open questions about the use of evidence in DFID, including about the quality of evidence used in decision-making and whether or not evidence-based decision-making is valued in the Department.

The survey ended with a vignette, asking the respondent to imagine that they were assessing options for a new DFID programme in the field of education. They are explicitly informed that the programme’s objective is to improve learning outcomes. They are told that they have commissioned research in the place in which the programme will be run, performing an A/B comparison of two kinds of intervention by means of a randomised control trial (RCT), and given the results of this RCT. They are then asked to simply choose which of the two interventions they would select, noting that they will personally sign the submission of the business case to the Secretary of State. In the control group, this is the extent of the problem. The full text of the vignette is as follows:

“You are responsible for a business case for a new GBP35 million education programme with the objective of improving student learning outcomes and will sign the covering submission requesting approval by the Secretary of State. Learning outcomes are a ministerial priority and manifesto commitment.

In preparation for your new business case, you commissioned a study that provides strong, experimental evidence in a number of regions of the country you work in on the long-term effects of both curriculum development and teacher performance incentives. The results of this study are summarised below. (For logistical reasons, the total number
of students in each of the two groups is not exactly the same, but this does not prevent the assessment of the results.)

<table>
<thead>
<tr>
<th></th>
<th>Students with improved scores</th>
<th>Students whose scores show no change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curriculum Development</td>
<td>2230</td>
<td>750</td>
</tr>
<tr>
<td>Teacher incentives</td>
<td>1070</td>
<td>210</td>
</tr>
</tbody>
</table>

In light of the information provided, which of the following recommendations would you make in your business case and covering submission:

1. To use the programme to improve curriculum development across the country.
2. To use the programme to roll out a system of teacher performance incentives across the country.”

It is worth making explicit a few points about the construction of this vignette. Firstly, selecting between the two options is simply a matter of comparing the ratio of successful outcomes (students with improved scores) to unsuccessful outcomes (students whose scores show no change). In the example above, the ratio for curriculum development is about 3:1 and that for teacher incentives is about 5:1. We randomised which of the two interventions was better, so some respondents would see teacher incentives with the better ratio, and others would see curriculum development with the better ratio. The numerical skills required to successfully compute this are well below the threshold even of a generalist civil servant. Anyone who had successfully completed the recruitment process outlined above should be able to calculate and compare these ratios.

Secondly, the sentence “learning outcomes are a ministerial priority and manifesto commitment” is of particular significance to civil servants. Manifesto commitments are promises made by the Government in power while campaigning for election, and it is considered an extremely serious matter to break them. They are the collective responsibilities of the Government. They supersede
the preferences of any individual Minister, since all Ministers campaign on the basis of the Manifesto, and if elected are expected to adhere to these promises.

Thirdly, the size of the business case, GBP35 million was carefully chosen. It is a large business case and as such respondents should infer that the (hypothetical) stakes are high. However, it falls just below the threshold for independent scrutiny by the Department’s Quality Assurance Unit, which is set at GBP40 million. This means that while a decision must be taken carefully (which is of course true no matter what the size of the business case), there is a relatively low chance of external scrutiny on the decision taken by the respondent.

Finally, the fact that the respondent is told that they will sign the covering submission that presents the final business case to the Minister indicates that they will be personally identifiable as the author of the decision taken.

The control treatments were identical except for the addition of a single new paragraph, providing information on Ministerial preferences. The additional text read as follows:

“The Secretary of State has been giving a series of interviews in which she has stressed the importance of curriculum development [OR teacher incentives] for improving learning outcomes, which she has said is one of her key objectives.”

This paragraph was placed directly after the first paragraph of the control treatment text. The paragraph provides information on what the most senior Minister in the department has publically announced as her preferred method of achieving learning improvements. However, it also reiterates that achieving success in this area is one of her key objectives.
Using this additional text, I created two treatment groups. Treatment group 1 were given a paragraph of information in which the Minister has publically proposed the less effective intervention. Treatment group 2 receive a paragraph in which the Minister has publically backed the more effective intervention.

The surveys were sent out using ‘Survey Monkey’ with a message following message requesting that staff fill out the survey, and a link to a video of the ‘survey champion’, a DFID senior manager, talking about the survey.
Annex B: Data and Attrition

Data

In addition to the vignette through which the survey experiment was implemented, the Evidence Survey collected a range of information useful as control variables and to explore causal pathways through additional analysis. The data collected included:

- Respondent grade (i.e. the level of seniority achieved in the Department)
- Respondent cadre (technical specialisation, if any)
- Respondent tenure (in DFID, a categorical variable, with three choices: less than five years, five to ten years and more than ten years)
- Gender
- Their self-assessed ability to interpret quantitative evidence
- Whether they were recruited locally (if international staff) or in the UK
- Their self-assessed level of comfort in challenging fellow civil servants
- Their self-assessed level of comfort in challenging Ministers
- Their confidence that standard DFID processes will pick up any errors in the use of evidence.

Though much of this information is valuable as control variables and for teasing out possible causal pathways, most of the data has the drawback of being categorical rather than continuous. This was unavoidable. Previous iterations of the Evidence Survey used such variables and the implementing team needed to ensure comparability. Choices have had to be taken in coding this categorical and ordinal data, and in particular, two should be highlighted.

---

19 Fully anonymised
20 The recruitment process for locally recruited staff is less intensive than for UK-recruited staff.
Firstly, I code civil service grade into a simple dummy variable, taking the value 1 if the respondent is a middle or senior manager and 0 if they are at any grade below this. One alternative to doing so would be to either treat grade as a continuous variable, coded from 0 to 6. This has the major drawback that there isn’t a clear linear relationship in power with grade. It is not clear that grade 2 is only half as powerful as grade 4; while a respondent at grade 6 (the ‘Senior Civil Service’) may have substantially more power than one at grade 5. Another alternative would be to split the grades into three strata, separating junior staff, middle managers and senior managers. Reassuringly, using any of these strategies preserves the main results of the paper, suggesting that data coding choices and functional form assumptions are not driving them.

Secondly, self-assessed ability to interpret quantitative evidence was also collected as a five-point likert scale, with 1 being no ability at all and 5 being extremely high ability. Again, there is no clear interpretation of this five point scale – is 2 twice as effective as 1, for example? What’s more, self-assessment is not particularly informative about differences in actual ability. Analysis of this variable suggests that there is a clear difference in those self assessing levels 4 and 5 and the rest, but no real difference between 4 and 5. As a result, I have chosen to code the information in this variable as a dummy for high ability, given the value 1 for those self-assessing their interpretive skills at 4 or 5 and 0 for the rest. This variable is used only as a control.

The experimental data collected was as follows:

- Dummy variables signifying assignation to treatment 1 (Minister back the less effective option) and treatment 2 (Minister backs the more effective option), with the reference category those assigned to the control group
• A 0/1 dummy variable for whether or not the respondent selects the more effective option in their business case.

Sampling and attrition

We sampled 989 of DFID’s staff to survey. The sampling strategy followed a stratified randomisation design, to match the previous Evidence Survey in 2013. There was no possibility of changing this sampling strategy, as consistency and comparability was a key requirement for the team implementing the survey.

The population of the survey was from all DFID staff of B2 grades and above (i.e. those likely to make some programming or policy decisions and use evidence in their work), across DFID’s two headquarters (one on Whitehall in London and the other in Abercrombie House in Scotland), and its Country Offices. A complete list of this population was stratified by grade (B2 – B1 – B1D – A2L – A2 – A1 – SCS) and further sorted by professional specialism. The number in each stratum was calculated based on the proportion of the relevant cadre and grade in the overall population. Some non-advisory roles and political positions were filtered out of the survey. The survey sample was randomly selected from each stratum, and then assigned randomly to treatments 1, 2 or control groups (without any stratification) and sent the survey link to complete within a 6-week deadline.

We sent two reminders and used a publicity campaign using internal communications and media to boost response rate.
Despite these efforts, our response rate was only 61% (604 individuals) - high for typical civil service surveys but lower than ideal for research purposes. At this stage attritors will not have seen any part of the survey and will be unaware of the existence of the experimental vignette, let alone to which treatment group they are assigned: we are simply measuring whether or not they ever even opened the survey instrument (attrition once the survey was opened is treated separately, below). Nevertheless it is possible that certain kinds of civil servant were systematically less likely to respond to the survey, which would compromise external validity – the extent to which the results presented here can be generalised to the entire population of civil servants in DFID.

To see if this was the case, I ran a regression to see if any respondent characteristics predicted their likelihood of opening the survey at all. Again using the linear probability model on a binary ‘opened survey/did not open survey’ dependent variable I find that few respondent characteristics are predictive of whether or not the survey is opened. Table B1 summarises. (The data set from which sampling was undertaken was more detailed than that collected by the survey – as a result, we have data on tenure in DFID which is precise to the year for the analysis of response rates, but not for the survey experiment).

<table>
<thead>
<tr>
<th>Table B1: Survey participation regressed against respondent characteristics. Dependent variable is a dummy for survey participation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable: Opened survey</td>
</tr>
<tr>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Tenure</td>
</tr>
<tr>
<td>Grade</td>
</tr>
<tr>
<td>Role</td>
</tr>
<tr>
<td>Senior Management</td>
</tr>
<tr>
<td>Policy Professionals</td>
</tr>
<tr>
<td>Programme Managers</td>
</tr>
<tr>
<td>Category</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Econ, Evaluation and Statistics</td>
</tr>
<tr>
<td>Human Development</td>
</tr>
<tr>
<td>Governance and Conflict</td>
</tr>
<tr>
<td>Climate, Humanitarian and Livelihoods</td>
</tr>
<tr>
<td>Private Sector and Infrastructure</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Directorate</th>
<th>B</th>
<th>SE</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country Programmes</td>
<td>.0761041</td>
<td>.100542</td>
<td>0.449</td>
</tr>
<tr>
<td>Economic Development</td>
<td>.0198882</td>
<td>.1035307</td>
<td>0.848</td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>.0993626</td>
<td>.1079949</td>
<td>0.358</td>
</tr>
<tr>
<td>Policy and Global Issues</td>
<td>.1217193</td>
<td>.0940114</td>
<td>0.196</td>
</tr>
<tr>
<td>Top Management Group</td>
<td>.0477121</td>
<td>.1762295</td>
<td>0.787</td>
</tr>
</tbody>
</table>

| HQ Dummy                                      | .0503626 | .0477455 | 0.292 |
| Constant                                      | .6535393*** | .0869407 | 0.000 |

* Statistically significant at the 10% level ** Statistically significant at the 5% level *** Statistically significant at the 1% level

Discussion of these results is in the main text of the paper.
REFERENCES

DFID. 2016. Smart Rules for Better Programme Delivery.


