How do merit commissions affect judicial behavior? Evidence from the Court of Justice of the European Union

Sivaram Cheruvu[1]   Joshua C. Fjelstul[2]

Silje Synnøve Lyder Hermansen[3]   Daniel Naurin[4]

Abstract

Governments create monitoring entities to ensure that policymakers are working effectively. Increasing transparency in the retention and selection of policymakers, however, may create incentives that lead to sub-optimal outcomes. One such entity is a "merit commission" (also known as a judicial council) that evaluates judges subject to retention and/or selection. How do merit commissions affect judicial behavior? Using a formal model, we show that lower-competence incumbent judges only subject to merit retention will complete cases more quickly at the expense of quality opinion writing, as their productivity is a signal of their effort. Conversely, judges subject to merit selection and subsequent retention will not make the same tradeoff, as it is more costly for a commission to negatively evaluate a judge upon retention after it has already selected them. We provide evidence leveraging the Article 255 panel for (re)appointments to the Court of Justice of the European Union.

[1]Department of Economics, Political & Policy Sciences, University of Texas at Dallas, sivaram.cheruvu@utdallas.edu, sivaramcheruvu.com.

[2]Department of Political Science and International Relations, University of Geneva, ARENA Centre for European Studies, University of Oslo, joshua.fjelstul@unige.ch, joshuafjelstul.com.

[3]Department of Political Science, University of Copenhagen, silje.hermansen@ifs.ku.dk, siljehermansen.github.io.

[4]ARENA Centre for European Studies, University of Oslo daniel.naurin@arena.uio.no.

Introduction

Policy-making is an inherently difficult task. To alleviate this burden, governments divide policy-making responsibilities among different institutions. This division of labor, however, creates the practical concern of whether these institutions are properly fulfilling their policy-making mandates. One solution to this problem is to subject these institutions to monitoring by another entity that can sanction ineffective policymakers. Depending on the conditions, such monitoring can lead to sub-optimal outcomes (e.g., Bueno de Mesquita and Stephenson 2007). The monitor, for example, may reward more productive policymakers (i.e., those that complete more tasks). For policymakers that are not sufficiently competent, the pressure to be more productive may lead to policy of lower quality. In this paper, we analyze two monitoring functions: *ex post* retention and *ex ante* selection of policymakers. We argue policymakers only subject to *ex post* retention will prioritize productivity to avoid sanctioning from the monitor. In contrast, policymakers subject to *ex ante* selection and then retention will not feel as much pressure to be more productive, as the monitor's evaluation of them as fit to serve on the institution prior to their appointment makes it more costly for the monitor to sanction them upon retention.

The selection and retention of judges on courts illustrates this tension. A wide variety of mechanisms exist to select and retain judges including judicial elections, executive appointment, and legislative appointment (e.g., Tiede 2023; Voeten 2007). Given appointers' desire across all such systems to make informed appointment decisions, many governments have created merit commissions.[1] Such commissions evaluate whether a judge is suitable to serve on a court. Following this evaluation, the relevant appointing institution decides whether to appoint the judge to the court.

---

[1] Tiede (2023) finds that a majority (68 out of 125) of the supreme courts she investigated involves a judicial council in the selection procedure. In New Mexico, a merit commission selects candidates that then immediately stand for partisan election (Canes-Wrone, Clark and Park 2012, 212). Merit commissions in South Carolina and Virginia select candidates that are then subject to legislative appointment (e.g., Tobias 2008).

In many civil law states, merit commissions — also known as judicial councils (e.g., Garoupa and Ginsburg 2009) — often have the power to retain and promote judges. A recent review of judicial selection procedures notes that further study of judicial councils "is much needed as their use is on the rise across the world" (Tiede 2023, 23). While scholars have evaluated how the presence of a merit commission affects judicial decision-making relative to other selection and retention mechanisms (e.g., Canes-Wrone, Clark and Kelly 2014; Gordon and Huber 2007), they have seldom considered how merit selection of a judge affects judicial decision-making in comparison to merit retention of a judge.

We argue that merit retention should affect judicial decision-making differently than merit selection. To provide the micro-foundations of our argument, we formalize an account of judges' effort investments as a strategic decision to signal their competence to a merit commission. Less competent judges *only* subject to merit retention will complete cases more quickly to signal to the merit commission that they are a competent judge deserving of retention. Since these judges have lower competence, however, expending more effort to complete cases more quickly leads to lower quality opinion-writing. Judges subject to merit selection will not make the same tradeoff, as the commission's selection of them prior to their initial appointment makes it more costly for the commission to evaluate them negatively upon retention. Put simply, if the commission provides a negative evaluation upon retention to a judge that they selected in the first place, it amounts to an admission that the initial selection was a mistake. As a result, relative to judges *only* subject to merit retention, judges subject to merit selection will complete cases more slowly, leading to higher quality opinion-writing.

To evaluate our hypotheses, we leverage the Article 255 Panel (255 Panel) for appointments to the two bodies of the Court of Justice of the European Union (CJEU): the Court of Justice (CJ), and the General Court (GC). The EU's Lisbon Treaty charged the 255 Panel with providing a positive or negative evaluation of incumbent judges subject to reappointment and new judges member states selected to serve their first term on the CJEU. The 255 Panel's creation in 2010 provides a unique opportunity: incumbent judges sitting on the

CJEU prior to 2010 were subject to an evaluation upon retention. The 255 panel selected — i.e., gave a positive evaluation prior to appointment — all judges that started their term after 2010. When comparing judges *only* subject to retention compared to judges subject to selection and then retention, we find judges *only* subject to retention at the GC wrote their judgments more quickly and less thoroughly, as the 255 panel's beliefs about their competence is lower relative to the CJ. Judges subject to selection and then retention at both the CJ and GC wrote their judgments more slowly, providing evidence for our hypotheses.

We organize our paper as follows. First, we examine the existing literature on judicial selection and retention, formally theorize over the process of merit selection and retention, and derive three testable hypothesis. Second, we describe the origins and operation of the 255 Panel for (re)appointments to the CJEU. Third, we provide our data and relevant model specifications to evaluate our hypotheses. Fourth, we present our results, which support our hypotheses. Lastly, we conclude by discussing our article's implications for the broader judicial politics literature.

Selection, Retention, and Judicial Decision-making

The procedures for selection and retention can directly impact both the composition of the bench and how judges decide cases in the courtroom. Indeed, these choices in the design of judicial institutions attempt to fill seats with highly qualified judges and to balance insulating judges from external pressures with the norms of accountability in liberal democracies. In the American context, scholars have examined the effects of partisan, non-partisan, and non-competitive retention elections on a variety of outcomes including decisions in abortion cases, death penalty cases, and general criminal sentencing (e.g., Choi, Gulati and Posner 2010; Kritzer 2016). Similarly, scholars have examined the differences in partisan and non-partisan selection on the composition of the judiciary (e.g., Bonica and Sen 2017). Institutional

3

retention and selection procedures, thus, create incentives affecting judicial behavior in the face of retention and the types of judges appointers select to sit on a court.

Some such procedures involve the use of a merit commission — an independent commission made up of some combination of lawyers, academics, and bureaucrats — that has the power to evaluate the qualifications of a candidate to sit on a court. The candidate then faces retention at the expiration of their term. We refer to such systems as *merit selection*, irrespective of the body responsible for retention of the judge. The appeal of such a system is straightforward. By removing judges from the plebescitary pressures inherent in direct election or direct appointments, while also subjecting them to a rigorous evaluation, judges are incentivized to perform by producing high quality decisions in time, while also being insulated from political pressure. For these ideals, numerous legal organizations – such as the American Bar Association and the State Bar Association (e.g., Goelzhauser 2018) – and scholars (e.g., Devaney 2019) have advocated for the merit selection of judges.

Whether merit selection lives up to these ideals is subject to debate. Studying state and federal judiciaries, Bonica and Sen (2017) find that the political ideology is less salient when judges are selected via merit selection committees. Goelzhauser (2018), in a study of Nebraska, provides evidence that merit selection committees evaluate legal merits differently than the appointing governor, but that partisan considerations are present also in merit commissions. With regards to judicial behavior, Owens et al. (2015) find no effect of judicial selection methods on the likelihood that state judges' decisions are appealed or reversed in the U.S. Supreme Court. Canes-Wrone, Clark and Kelly (2014), additionally, find that state supreme court justices subject to merit selection are not insulated from plebiscitary pressures when deciding to uphold death penalty decisions. Choi, Gulati and Posner (2010) provide evidence that merit selected judges are more productive (in terms of writing many opinions) than appointed judges, but less productive than elected judges. Furthermore, the merits-selected judges' opinions were cited less often than appointed judges' opinions and demonstrated less independence as measured by opinions against copartisans.

Outside of the United States, merit commissions — more commonly known as judicial councils — are also involved in the retention (and promotion) of judges. As Garoupa and Ginsburg (2009, 114) explain, "Merit Commissions can be seen as analogous to judicial councils [...] Because in common law systems, the judiciary is not a 'career judiciary' in the civil law sense, there is less interest in having independent commissions handle discipline, promotions, and reassignments, and greater emphasis on initial appointments." In a comparative cross-country study of judicial councils, Garoupa and Ginsburg (2009) fail to find any systematic relationship between the composition of judicial councils (in terms of the number of judges being members of the council) or their delegated powers (in terms of competences regarding the selection and monitoring of judges) and comparative indexes of judicial independence and quality.

In many civil law countries, law graduates must complete some form of public examination to accede to the judiciary and usually do not need to have experience. Entering the judiciary, thus, does not involve a rigorous evaluation by a merit commission. Nonetheless, the career pressures inherent in retention can affect judicial decision-making, as many councils involve senior judges in the judicial hierarchy or bureaucrats with political interests making retention decisions (e.g., Ramseyer and Rasmusen 2003). Although judges in these settings generally are not subject to renewable terms, we conceptualize retention broadly as also encompassing a commission's decision to renew, promote, or discipline a judge with regards to their current posting. We refer to such systems as *merit retention*.

Given this division in how merit commissions operate across contexts, scholars have seldom examined how and whether the mechanisms of merit selection and merit retention affect judicial decision-making differently. To provide the micro-foundations of these mechanisms, we present a formal model of the interaction between a merit commission and a judge seeking reappointment. Our model focuses on reappointment, rather than the decision of a commission to recommend a new judge for a first term, as all judges in systems with renewable terms or periodic evaluation by a judicial council face retention irrespective of whether they were

subject to merit selection. Specifically, we examine how the incentives the merit commission's evaluation generates affects the output of judges who need a favorable recommendation to retain their seat on the court.

Formal Model

The objective of the merit commission is to give favorable recommendations to sufficiently competent judges and to give unfavorable recommendations to insufficiently competent judges. However, the merit commission cannot observe the competence of the judge directly. It would be infeasible for the commission to thoroughly evaluate the entire judicial output of a judge, given the large number of cases that judges often hear (e.g., Bielen et al. 2018). As such, merit commissions evaluate judges based on observable criteria — metrics on which sufficiently competent judges will perform well and insufficiently competent judges will not.

Our model captures the fundamental tension in the interaction between merit commissions and judges: a merit commission does not know the competence of a judge, although it has a belief about it, and bases its recommendation on the level of effort that judges expend to satisfy the reappointment criteria it establishes. Lower-competence judges have an incentive to mimic the level of effort expended by higher-competence judges in order to get a favorable recommendation. Therefore, to separate themselves from lower-competence judges, higher-competence judges have to expend a high enough level of effort to satisfy the commission's reappointment criteria that lower-competence judges, for whom expending effort is more costly, are unwilling to expend.

Structure and Strategies

The model has two players: a judge, J, and a merit commission, C. The judge has a type, ω, which represents the competence of the judge. Figure 1 shows the order of play, and Table 1
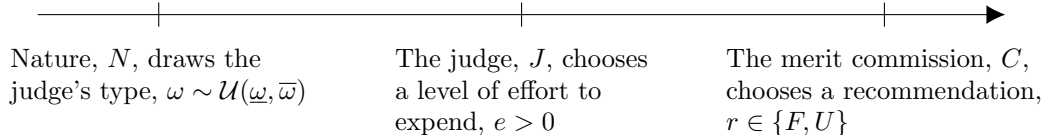
6

Figure 1. The order of play.

Note: This figure shows the order of play. The game has two choice variables: the level of effort that the judge chooses to expend and the recommendation of the merit commission.

summarizes the model's notation. Nature, N, starts the game by choosing the judge's type, ω. We assume that the judge's type is uniformly distributed between a lower bound and an upper bound, $\omega \sim \mathcal{U}(\underline{\omega}, \overline{\omega})$, where $\underline{\omega} > 0$ and $\overline{\omega} > \underline{\omega}$. Note that changing $\overline{\omega}$, relative to $\underline{\omega}$, manipulates the expected competence of the judge, $\mathbb{E}[\omega] = \frac{\overline{\omega} + \underline{\omega}}{2}$.

The judge's type is private information. The judge knows their own type as soon as Nature draws it, but the merit commission does not know their type until after the game has ended. However, the commission does know the distribution of the judge's type — that is, the players share a common understanding of the range of possibilities. The merit commission's uncertainty over the type of the judge is the central tension in the model. As we will see, the commission wants to give a favorable recommendation to sufficiently high types, representing high-competence judges, and an unfavorable recommendation to insufficiently high types, representing a low-competence judges.

After Nature selects the judge's type, the judge chooses a level of effort, $e > 0$, to expend on satisfying the criteria of the merit commission. The commission cannot directly observe the judge's type, but it can directly observe the level of effort that the judge expends. After observing the level of effort expended by the judge, the commission chooses a recommendation, $r \in \{F, U\}$. It can choose to give a favorable recommendation, F, or an unfavorable recommendation, U. The recommendation ends the game.

A strategy profile for a player proscribes an action for the player in response to all possible actions by the other player, regardless of whether the other player would actually play them in equilibrium. In equilibrium, each player's strategy profile is a best response to the other

Table 1. Model Notation

Random variables	
$\omega \sim \mathcal{U}(\underline{\omega}, \overline{\omega})$	The type of the judge, which represents their competence. This parameter is uniformly distributed between $\underline{\omega}$ and $\overline{\omega}$. It is private information to the judge.
Choice variables	
$e > 0$	The level of effort expended by the judge in order to convince the merit commission to give a favorable recommendation.
$r \in \{F, U\}$	The recommendation of the merit commission, where F is a favorable recommendation and U is an unfavorable recommendation.
Exogenous parameters	
$\underline{\omega} > 0$	The lower bound of the type distribution.
$\overline{\omega} > \underline{\omega}$	The upper bound of the type distribution. Changing this parameter relative to $\underline{\omega}$ changes the expected competence of the judge, $\mathbb{E}[\omega]$.
$b > 0$	The expected career benefit to the judge of receiving a favorable recommendation from the merit commission.
$c > 0$	The expected political cost to the merit commission for giving an unfavorable recommendation.
$\omega_i \in (\underline{\omega}, \overline{\omega})$	The minimum standard of competence that is acceptable to the merit commission. When $\omega > \omega_i$, giving a favorable recommendation is beneficial to the commission. Otherwise, it is costly.
Endogenous parameters	
$\underline{e}^* > 0$	The weak signal in equilibrium, sent by types $\omega < \omega^*$.
$\overline{e}^* > \underline{e}^*$	The strong signal in equilibrium, sent by types $\omega \geq \omega^*$.
$\omega^* \in (\underline{\omega}, \overline{\omega})$	The cut-point on the ω that determines whether a type plays the weak signal or strong signal in equilibrium.

player's strategy profile. A strategy profile for the judge, $\sigma_{J|\omega}(\sigma_C)$, is a function that maps a level of effort, $e > 0$, conditional on their type, ω, to the action of the merit commission, $r \in \{F, U\}$. A strategy profile for the merit commission, $\sigma_C(\sigma_{J|\omega})$, is a function that maps a recommendation, $r \in \{F, U\}$, to the action of the judge, e.

Preferences

We make several reasonable assumptions about the preferences of the judge and the merit commission. First, we assume that the judge is career-motivated and wants to be reappointed to the court (e.g., Epstein, Landes and Posner 2011). The judge needs a favorable recommendation from the merit commission in order to be reappointed. Thus, the judge receives an expected personal benefit for getting a favorable recommendation.

Second, we assume that merit commissions want to give favorable recommendations to sufficiently competent judges and unfavorable recommendations to insufficiently competent judges. The merit commission has a minimum standard of competence, which is known by the judge. If the commission gives a favorable recommendation to a judge that meets or exceeds this standard, it receives a benefit. However, it incurs a cost if it gives a favorable recommendation to a judge that falls short of this standard.

Third, we assume that expending effort to convince the merit commission to give a favorable recommendation is costly for the judge. Specifically, the cost of effort is directly proportional to the level of effort expended: a one-unit increase in effort is a one-unit increase in the cost of effort. This functional form is not critical to the equilibrium we analyze, but provides analytical tractability.

Fourth, we assume that expending effort is less costly for high-competence judges than for low-competence judges. In other words, the type of the judge influences the marginal cost of effort. In fact, this ability is precisely what makes a judge competent — they can expend more effort at a lower cost.

Fifth, we assume that it is politically costly for the merit commission to issue a negative recommendation. Doing so is likely to embarrass the government or politician who recommended or appointed the judge and risks upsetting other political actors who support the judge. One reason a negative recommendation could be costly is that members of a merit commission have their own career concerns that could be harmed by rejecting a judge.

Based on these assumptions, we specify utility functions for both players. If the merit commission chooses to give a favorable recommendation, the judge receives a benefit, b. Regardless of whether the commission gives them a favorable recommendation, the judge pays a cost proportional to the level of effort they choose to expend, $-\frac{e}{\omega}$, and the marginal cost is smaller for more competent judges.

$$u_{J|\omega}(F \mid e) = b - \frac{e}{\omega}$$

$$u_{J|\omega}(U \mid e) = -\frac{e}{\omega}$$

If the merit commission gives a favorable recommendation, the commission receives a cost or benefit, $\omega - \omega_i$, depending on how competent the judge is. The commission does not know the type of the judge, but it has a belief about it. If the type of the judge exceeds the minimum standard of competence expected by the commission, $\omega > \omega_i$, then the commission receives a benefit for giving a favorable recommendation. However, if the type of the judge falls short of that standard, $\omega < \omega_i$, the commission pays a cost. If the commission gives an unfavorable recommendation, it pays a political cost, $-c$, for calling out the government that appointed the judge. This cost would be higher if the merit commission was involved in selecting the judge, as giving an unfavorable recommendation is then also an implicit admission that the commission erred in its initial selection of the judge.

$$u_C(F \mid \omega) = \omega - \omega_i$$

$$u_C(U \mid \omega) = -c$$

These preferences capture the fundamental tension in the model. The merit commission is uncertain about the competence of the judge — that is, it does not precisely know the value of ω — and has to weigh the risk of giving a favorable recommendation to an insufficiently competent judge against the political cost of rejecting the judge. The judge wants to expend

just enough effort to convince the commission to give a favorable recommendation, and no more, as effort is costly.

Equilibrium Analysis

We derive a unique semi-separating perfect Bayesian equilibrium (PBE) in cut-points with two on-path equilibrium signals. Types below a cut-point send a weak signal and types above the cut-point send a strong signal. The merit commission gives a favorable or unfavorable recommendation based on the signal that it observes. See Supporting Information Section B for a more technical presentation of the equilibrium and formal proofs.

Proposition 1. There exists a unique semi-separating perfect Bayesian equilibrium (PBE) in cut-points with two on-path equilibrium signals. This equilibrium has the following properties:

1. **Strategy of the judge:** Types above a cut-point, $\omega \geq \omega^*$, send a strong signal, $\overline{e}^*$, and types below the cut-point, $\omega < \omega^*$, send a weak signal, $\underline{e}^*$.

2. **Strategy of the merit commission:** When the commission observes a sufficiently strong signal, $e \geq \overline{e}^*$, it gives a favorable recommendation. Otherwise, $e < \overline{e}^*$, it gives an unfavorable recommendation.

3. **On-path beliefs of the merit commission:** When the commission observes the strong signal, $\overline{e}^*$, it believes $\omega \sim \mathcal{U}(\omega^*, \overline{\omega})$. When it observes the weak signal, $\underline{e}^* = 0$, it believes $\omega \sim \mathcal{U}(\underline{\omega}, \omega^*)$.

4. **Off-path beliefs of the merit commission:** When the commission observes an unexpectedly weak signal, $\underline{e}^* < e' < \overline{e}^*$, it believes $\omega \sim \mathcal{U}(\underline{\omega}', \overline{\omega}')$. When it observes an unexpectedly strong signal, $e'' > \overline{e}^*$, it believes $\omega \sim \mathcal{U}(\underline{\omega}'', \overline{\omega}'')$.

5. **Equilibrium cut-points:** The strong signal is $\overline{e}^* = b\big(2(\omega_i - c) - \overline{\omega}\big)$, the weak signal is $\underline{e}^* = 0$, and the cut-point on ω is $\omega^* = 2(\omega_i - c) - \overline{\omega}$.

11

6. **Equilibrium existence:** This semi-separating equilibrium exists if and only if (a) the expected competence of the judge is sufficiently low, $\overline{\omega} < 2(\omega_i - c) - \underline{\omega}$, and (b) the commission believes that an unexpectedly high signal is more likely to come from a higher-competence judge and an unexpectedly low signal is more likely to come from a lower-competence judge, $\underline{\omega}'' + \overline{\omega}'' > \underline{\omega}' + \overline{\omega}'$.

In equilibrium, sufficiently competent judges, $\omega \geq \omega^*$, send a strong signal by expending a high level of effort, $\overline{e}^*$, and insufficiently competent judges, $\omega < \omega^*$, send a weak signal by expending a low level of effort, $\underline{e}^* = 0$. The merit commission gives a favorable recommendation when it sees a sufficiently high level of effort, $e \geq \overline{e}^*$, and gives an unfavorable recommendation otherwise.

In this equilibrium, the merit commission has on-path beliefs and off-path beliefs. The commission's on-path beliefs are the commission's beliefs when it sees a signal that the judge is supposed to send in equilibrium, which is either the strong signal, $\overline{e}^*$, or the weak signal, $\underline{e}^*$. The commission's off-path beliefs are its beliefs when it sees an unexpected signal, which could be any $0 < e' < \overline{e}^*$ or $e'' > \overline{e}^*$.

The commission's prior belief — its belief before it observes the judge's signal — is that the judge's type, ω, is uniformly distributed between a lower bound, $\underline{\omega}$, and an upper bound, $\overline{\omega}$: $\omega \sim \mathcal{U}(\underline{\omega}, \overline{\omega})$. The commission then updates its beliefs based on the judge's signal, and that change in beliefs determines whether or not the commission is willing to give a favorable recommendation. The strong signal, $\overline{e}^*$, leads the commission to give a favorable recommendation, and the weak signal, $\underline{e}^* = 0$, leads the commission to give an unfavorable recommendation.

Observing the strong signal, $\overline{e}^*$, causes the commission to believe that the type of the judge, ω, is uniformly distributed between a cut-point, ω^*, and the upper bound, $\overline{\omega}$: $\omega \sim \mathcal{U}(\omega^*, \overline{\omega})$. Thus, the commission believes that the judge is more competent than it did before, convincing it to give a favorable recommendation. Observing the weak signal, $\underline{e}^* = 0$, on the other hand, causes the commission to believe that the type of the judge, ω, is uniformly

distributed between the lower bound, $\underline{\omega}$, and a cut-point, ω^*: $\omega \sim \mathcal{U}(\omega^*, \overline{\omega})$. Thus, the commission believes that the judge is less competent than it did before, leading it to give an unfavorable recommendation.

This equilibrium is semi-separating in the sense that some lower-competence types, who fall short of the commission's minimum standard of competence, $\omega < \omega_i$, mimic higher-competence types, who do satisfy the standard, $\omega \geq \omega_i$. As long as sufficiently few lower-competence types are willing to do this, the commission is still willing to risk a positive recommendation when it sees the strong signal, $\overline{e}^*$. The level of effort expended by the judge, $\overline{e}^*$, has to be high enough that enough lower-competence judges are not willing to send that signal. The fact that the marginal cost of effort is more costly for lower-competence judges makes this possible.

There are two conditions that have to be satisfied for this semi-separating equilibrium to exist. First, the expected competence of the judge has to be sufficiently low.[2] If the judge is sufficiently competent in expectation, the merit commission will always prefer to give a favorable recommendation because the risk of giving a favorable recommendation to an insufficiently competent judge is low compared to the expected political cost of giving a negative recommendation.

Second, the commission has to believe that an unexpectedly weak signal, $0 < e < \underline{e}^*$, is more likely to come from a lower-competence type and an unexpectedly strong signal, $e'' > \overline{e}^*$, is more likely to come from a higher-competence type. In other words, the commission needs to believe that the judge is sufficiently competent when it sees a higher-than-expected level of effort and that the judge is insufficiently competent when it sees a lower-than-expected level of effort.

These off-path beliefs are plausible. Expending effort to satisfy the reappointment criteria of the merit commission is always costly, and if the judge does make a mistake, and sends

---

[2] Specifically, $\overline{\omega} < 2(\omega_i - c) - \underline{\omega}$ (see Proposition 1). Recall that changing $\overline{\omega}$ manipulates the expected competence of the judge, $\mathbb{E}[\omega] = \frac{\overline{\omega} + \underline{\omega}}{2}$.

an unexpected signal, they should be more likely to make low-cost mistakes than high-cost mistakes. Consequently, it makes sense for the commission to believe that stronger-than-expected signal, $e'' > \overline{e}^*$, is more likely to come from a relatively high-competence type than a relatively low-competence type, as such a mistake is relatively more costly for the lower-competence type. This is especially true because the marginal cost of effort is higher for lower-competence types. In addition, it makes sense for the commission to believe that a lower-competence judge would be more likely to make a mistake that would cause them not to be reappointed, $0 < e < \underline{e}^*$, than a higher-competence judge.

Comparative statics

We derive comparative statics to identify the conditions under which the level of effort a judge has to expend in equilibrium on satisfying the reappointment criteria of the merit commission, in order to get a favorable recommendation from the merit commission, is higher or lower. This speaks to the question of when merit commissions will influence the behavior of judges on the bench to a greater or lesser degree.

Looking at Proposition 1, the level of effort that convinces the merit commission to give a favorable recommendation in equilibrium, $\overline{e}^*$, depends on four exogenous parameters: (1) the benefit to the judge of reappointment, b; (2) the political cost to the merit commission of giving an unfavorable recommendation, c; (3) the merit commission's minimum standard of competence, ω_i; and (4) the upper bound on the distribution of the judge's type, $\overline{\omega}$, which manipulates the judge's competence in expectation. In this section, we analyze how a change in each of these exogenous parameters affects the signal that the judge has to send in order to convince the merit commission to give a positive recommendation, $\overline{e}^*$. See Supporting Information Section B for formal proofs.

Note that we cannot derive testable predictions over the two other endogenous parameters in the model, $\underline{e}^*$ and ω^*. The weak signal in equilibrium is a constant, $\underline{e}^* = 0$, as expending effort is always costly for the judge, and therefore does not change with respect to the

14

exogenous parameters. Three of these exogenous parameters also affect the cut-point on the judge's type in equilibrium, ω^*, and therefore which types send which signal. Furthermore, we cannot measure the cut-point empirically, unlike the level of effort that the judge expends, e^*.

Result 1. In equilibrium, the level of effort expended by the judge to satisfy the criteria of the merit commission, $\overline{e}^*$, is increasing in the benefit to the judge of being reappointed, b. As b increases, $\overline{e}^*$ increases.

Naturally, when the reappointment benefit increases, the judge expends more effort to convince the merit commission they are sufficiently competent. When the reappointment benefit increases, lower-competence judges have a greater incentive to mimic high-competence judges by sending the strong signal, despite the cost of the effort required. For the merit commission to still be willing to give a favorable recommendation, despite its uncertainty over the type of the judge, the signal that the judge sends has to become more costly. This allows the signal to continue to separate out lower-competence judges despite those judges' increasing incentive to mimic higher-competence judges.

Result 2. In equilibrium, the level of effort expended by the judge to satisfy the criteria of the merit commission, $\overline{e}^*$, is increasing in the minimum standard of competence expected by the merit commission, ω_i. As ω_i increases, $\overline{e}^*$ increases.

As the merit commission's minimum standard of competence goes up, the judge has to expend more effort to get a favorable recommendation. When a merit commission demands a higher standard of competence, it no longer prefers to give a favorable recommendation to some lower-competence judges, as doing so becomes costly. To still be willing to give a judge a favorable recommendation, despite its uncertainty over the type of the judge, the judge has to convince the commission that they are sufficiently competent, according to the new standard. To separate themselves from those lower-competence judges, a judge has to

send an even stronger signal by expending a level of effort that would be too costly for the lower-competence types to mimic.

Result 3. In equilibrium, the level of effort expended by the judge to satisfy the criteria of the merit commission, $\overline{e}^*$, is decreasing in the cost of giving an unfavorable recommendation, c. As c increases, $\overline{e}^*$ decreases.

When giving an unfavorable recommendation is more costly, the commission is more willing to tolerate the risk of giving a favorable recommendation to some lower-quality judges who fall just below its minimum standard of competence. In other words, the merit commission is willing to tolerate less separation. This condition means that the level of effort a judge has expended in order to separate themselves goes down. From the judge's perspective, when giving an unfavorable recommendation is more politically costly to the commission, the threat of an unfavorable recommendation is less credible. A judge can therefore expend less effort and still expect to get a favorable recommendation.

Result 4. In equilibrium, the level of effort expended by the judge to satisfy the criteria of the merit commission, $\overline{e}^*$, is decreasing in the upper bound of the distribution of the judge's type, $\overline{\omega}$, which manipulates the expected competence of the judge, $\mathbb{E}[\omega]$. As $\overline{\omega}$, and therefore $\mathbb{E}[\omega]$, increases, $\overline{e}^*$ decreases. If $\overline{\omega}$ becomes too large, $\overline{\omega} > 2(\omega_i - c) - \underline{\omega}$, the semi-separating equilibrium collapses into a pooling equilibrium in which the commission always gives a favorable recommendation.

As the competence of the judge increases in expectation, the level of effort the judge has to expend goes down. Furthermore, if the expected competence of the judge is sufficiently high, the judge need not expend any effort at all, in which case the semi-separating equilibrium collapses into a pooling equilibrium where all judges send the same signal and the merit commission always gives a favorable recommendation. In other words, the merit commission will not impact the behavior of the judge.

In interpreting this comparative static, which is on the upper bound of the distribution of the judge's type, $\overline{\omega}$, it is important to note that increasing or decreasing $\overline{\omega}$ manipulates the expected value of their type, $\mathbb{E}[\omega] = \frac{\overline{\omega}+\underline{\omega}}{2}$. As such, we can interpret a change in $\overline{\omega}$ as a change in the competence of the judge in expectation, $\mathbb{E}[\omega]$. Substantively, the expected competence of the judge would be higher, for example, if the pool of candidates were more qualified or if the court were a higher court, compared to a lower court.

When the expected competence of the judge is increasing, the merit commission needs to be less careful about giving a favorable recommendation. More lower-competence judges can mimic sufficiently high-competence judges by sending the strong signal, and the commission will still be willing to give a favorable recommendation because the judge is sufficiently competent in expectation. Therefore, the judge can send a weaker signal and still convince the commission to do so.

If the expected competence of the judge becomes too high, however, the semi-separating equilibrium collapses into a pooling equilibrium in which all types send the same signal and the merit commission always gives a favorable recommendation. Recalling Proposition 1, the upper bound of distribution of the judge's type, $\overline{\omega}$, and therefore the expected value of the judge's type, $\mathbb{E}[\omega]$, has to be sufficiently small for a semi-separating equilibrium to exist.

Intuitively, if the competence of the judge is sufficiently high in expectation, the merit commission is always willing to give a favorable recommendation because the expected cost of giving a favorable recommendation to an insufficiently competent judge, $\omega - \omega_i$, is always preferable to the political cost of giving an unfavorable recommendation, $-c$. Thus, if the merit commission believes that the judge is sufficiently competent, the judge does not need to exert effort to signal their competence. This result demonstrates the limit of merit commissions: if judges are sufficiently competent, merit commissions will not impact their behavior on the bench.

17

Case selection and data

To test our predictions, we examine the introduction of a merit selection committee for appointments to the CJEU, the high court of the EU. The CJEU is well-suited for testing our theory. First, the reform came late in the Court's history, allowing comparison of screened and unscreened judges within the same institution. Second, the radical transparency in in the committee's assessment criteria means that judges can adjust their behavior to avoid deselection. It also means that we can measure their performance in similar ways to the committee itself. Third, the committee is nevertheless a critical test of the reputational cost implied in the reassessment of incumbent judges, since its conclusions are only advisory.

Governments use judicial appointments to influence the Court's decisions. Judges serve six-year renewable terms and have thus always undergone both screening and retention decisions informed by candidates' political affinity and ability to gain influence in office. The merit selection committee (the "Article 255 panel") did in no way disturb that logic (Hermansen and Naurin 2019). While some authors had hoped the reform would provide a check on governments' political influence (e.g., Pérez 2015), a more sober understanding considers the Article 255 panel as an agent advising on judges' competence (Sauvé 2015).

Governments care about judges' competence for two reasons. On the one hand, legal skills and diligence are common goods that increase the quality of the Court's decisions. Nominating governments may nevertheless face problems committing to such criteria. On the other hand, they may care about the quality of judges for purely instrumental reasons, but lack the information to make a satisfactory assessment. As long as judges' legal quality is instrumental to their influence, member states do not face a commitment problem. Influence in the Court is – at least partially – allocated according to judges' skills and and effort (e.g., Cheruvu 2019; Hermansen 2020). However, when that link is broken or influence is not easily observable, governments' heuristics fall short.

The Article 255 panel changed the informational environment in which appointments are made. Its unique goal is to assess judges' suitability for office (TFEU, art. 255) and is

18

in a position to acquire high-quality information that allows for a unified and transparent appraisal. We may expect, thus, the largest effect of merit selection in instances where governments otherwise lack the information or motivation to effectively assess their judges. Specifically, the Article 255 panel fulfills two informational purposes: i) It builds trust among member states by ensuring a mutual monitoring of *other* governments' nominees. ii) It helps governments make a better assessment of their *own* nominees.

The Article 255 panel aims at re-building governments' mutual trust (Sauvé 2015). Upon the panel's establishment, the Court's evolution in a federal direction had put judges' professional suitability to the fore. Yet member states had come to doubt the suitability of other governments' nominees. Member states considered merit selection for some time but they were also reluctant to yield their appointing authority. The final impetus seemingly came with the 2004 EU enlargement. Observers have in particular pointed to concerns about nominations from former authoritarian states (Dumbrovsky, Petkova and Van Der Sluis 2014).

Notwithstanding the enlargement – several former autocracies had also joined the EU in the 1980s – the reform came as the hitherto last step in the CJEU's evolution from a traditional international institution to the EU's de facto federal high court. By the early 2000s most domestic high courts had recognized the CJEU's competence to strike down government policies. Yet the transfer of sovereignty was no longer compensated by an intergovernmental check. Judges had initially reached most decisions in plenary, with appointees from all member states present. However, as the Court's case-load grew in the 1980s and 90s, more decisions where delegated to subsets of judges. While politically important, their decisions may not necessarily reflect the consensus that the full court would have reached (e.g., Fjelstul 2022). The future of disputed government policies was thereby in the hands of judges without the procedural guarantees that an intergovernmental setup ensures. The Nice Treaty (2003) finally formalized the current chamber system. Simultaneously to the

2004-enlargement, the CJEU thus moved in a federal direction, making the suitability of *other* member states' judges a pressing concern.

The Article 255 panel also helps governments screen their *own* judges. By the turn of the millennium another problem had surfaced. Governments were unable to effectively monitor their own judges. The reason can be traced back to the establishment of the lower-level General Court (GC) 15 years prior. It was set up to alleviate the Court of Justice's (CJ) workload and decrease processing time for cases of little political importance (Avril 2022). However, the difference in the types of cases and the internal organization between the two courts created an informational deficit and a commitment problem for GC nominations that the Article 255 panel alleviates.

Specifically, governments have used judges' responsibilities in office as cues of their influence over case law (Hermansen and Naurin 2019). The Court's internal organization allows rank-and-file judges to act as agenda setters (Hermansen 2020). Each case is managed by a reporting judge that collects information on behalf of their peers. They draft the first proposal for a decision, negotiate a consensus and write up the final judgment. Moreover, informal norms of policy specialization means that individual judges gain disproportionate influence over specific policy areas. For policy-oriented governments, the nomination of judges that acquire such positions is an opportunity to influence the Court.

However, the quality of those cues is lower for judges at the GC (Hermansen and Naurin 2019). First, there is a large variability in the legal complexity and political salience of cases in the CJ. This is reflected in a large variability in the share of influential decisions in judges' portfolio, providing valuable information to governments. By contrast, the GC mainly deals with run-of-the-mill cases. Second, cases are managed differently in the two courts. The upper-management at the CJ delegates responsibilities early in the process with little monitoring afterwards. Judges' personal characteristics are thus important to the quality of the judgment itself. Case allocation is therefore highly selective. Case allocation at the GC is, by

contrast, more mechanistic. Cases are allocated by the middle management, while the upper management supervises the judges' proposed solutions. GC judges are therefore treated as replaceable. The link between their competence and influence is thereby less apparent in the GC. This disturbs the informational cues that policy-oriented governments have relied on. As a result, while political orientation are important predictors for renomination in both courts, judges' performance is a criterion for merit retention only in the CJ (Hermansen and Naurin 2019).

The Article 255 panel re-actualized performance accountability for the GC. While governments had retained CJ judges for the relative *importance* of their case portfolio, the Article 255 panel focuses explicitly on judges' absolute *productivity*. The choice has a historical backdrop. The GC had been criticized for its backlog when the Article 255 panel was established. This could not be explained by the lack of judges. The EU enlargements had almost doubled the size of the Court, while the case load had increased only moderately. Instead, observers emphasized that productivity was low in parts due to the lack of performance accountability (Dehousse 2011, 7-8 and 16-17). Specifically, a high turnover among judges at the GC impeded on its normal functioning. Beyond delays following the reshuffling of dossiers, the Court had problems accumulating the experience required for quality judgments. The more challenging cases are often delegated to seasoned members (interview). Yet, in contrast to the CJ, hard-working judges were not shielded from deselection by their governments (Dehousse 2011, 7-8 and 16-17). The solution came in the form of closer monitoring and specific (de)selection criteria.

The GC adjusted its case-processing system in conjunction with the establishment of the Article 255 panel. Together, the two institutions reversed the trend already in the 2011-15 period (Dehousse 2016). Their collaboration means that the panel is better informed about judges' potential performance than governments.

Six of the its seven members are nominated by the President of the CJEU (the seventh member is nominated by the European Parliament). Among these are prominent former CJEU judges with first-hand experience of the skills required. The specificity of GC cases requires more substantial experience with EU law than in the CJ (Dehousse 2011, 7-8 and 16-17). Yet the legal competences of GC judges were said to vary substantially prior to the panel's establishment. Judges are assisted by law clerks, but deliberations take place in closed chambers where only members are admitted. There were rumors of judges whose French skills were insufficient to take part in the discussions, and of reporting judges with only summary knowledge of the content of their own proposals (interviews). Unsurprisingly, in its public communications, the panel emphasizes selection criteria such as legal capabilities, language skills and an ability to work in an international environment (Seventh activity report, 2022, p. 12). New nominees are interviewed in French with the aim *inter alia* to assess their knowledge of EU law. Participants have likened the experience to a second bar exam in which intimate knowledge of the CJEU and its case law is a clear asset (interview). The panel is thus well-suited to assess judges' legal competences, but it also stays in communication with the Court throughout judges' mandate.

In 2011, the GC defined milestones for each case, assorted with a target timeline. The list specifies for each judge the dates when cases are opened and closed, flagging all potential delays. While controversial, the "Tableau de productivité" is distributed to all judges, as well as to the Article 255 panel. In assessing judges for reappointment, the panel specifically "compares the duration of proceedings in the cases dealt with by candidates with the expected duration of proceedings in comparable cases, on the basis of the indicative internal deadlines of the court" (The seventh activity report, 2022, p. 14). While the panel has yet to refuse an incumbent candidate, it has stressed that under-performance according to such quantifiable criteria would justify a negative opinion.

The panel's pro-active approach to the screening of judges speaks to the credibility of this threat. As of now, 243 nominees have been assessed, resulting in 214 formal opinions of

which 28 have been unfavorable (The seventh activity report, 2022, p. 10), leading observers to claim the panel has overstepped its mandate (von Bogdandy and Krenn 2015, 173–74). Thus far, member states have followed the panel's formal recommendations. Moreover – reading between the lines (29 nominees are left unaccounted for) – we also understand that governments may retract nominations prior to the panel's official opinion; possibly saving aspiring judges, panelists and governments from the embarrassment of an invalidation.

The Article 255 panel's focus on quantifiable measures of productivity necessarily opens a discussion about a tradeoff with the quality of judges' work. However, measures of quality are harder to device (Dehousse 2011, 10). The Panel's comparison of the duration of proceedings in "comparable cases" alludes to a more substantive quality assessment. The research community often refers to the length of judgments and the number of citations as proxies for quality that the Court also strives to achieve.

Judgments are written following the French tradition for brief, deductive statements of a legal reasoning. The lack of inductive argument and polemic elaborations – common in the Anglo-American tradition – means that the Court's reasoning is hard to understand, discuss and adjust to. Efforts where done from the 1980s and onward to make the text more readable. Their coherence and legibility have particularly improved in recent years when the average length of preliminary reference judgments has increased. However the quality of judgment texts continues to vary substantially (De Waele 2021, 65–67). The Court also conveys how particular judgments are embedded in extant case law through its citations to previous decisions (e.g., Larsson et al. 2017). However, writing longer judgments and identifying relevant case law is time consuming (e.g., Lupu 2013), meaning judges' productivity may take a toll.

To test the implications of screening and merit retention on these tradeoffs, we draw on information about the reporting judges and the texts they prepare to construct indicators for effort similar to those used by the Court. Our data covers two decades (2000-2020) and

is drawn from the IUROPA CJEU Database Platform ([Brekke et al. 2023](#)) and its Text Corpus.

Hypotheses

We develop three testable hypotheses based on our comparative statics, focusing on the level of effort expended by judges eager to signal their competence to the merit commission, $\overline{e}^*$. We operationalize $\overline{e}^*$ using three measures: the duration of the case in days (DURATION), the length of the judgment in words (WORD COUNT), and the number of case law citations in the judgment (CITATIONS). The tradeoff between productivity and quality means that e^* is negatively correlated with DURATION, WORD COUNT, and CITATIONS. Note that this correlation means the predicted signs of the marginal effects in our hypotheses will be the opposite of the predicted signs in our comparative statics.

Our first hypothesis is based on Result 2, predicting that effort is increasing in the merit commission's minimum standard of competence, ω_i. Upon its creation, the Article 255 panel increased the standard for reappointment. Since effort corresponds to higher productivity and lower quality, we should therefore expect judges that were already in office when the standard incrased to work more quickly but produce lower-quality judgments (i.e. we compare RETENTION $= 1$ to RETENTION $= 0$), knowing that they now need to signal their competence to the panel.

> **Hypothesis 1.** Judges who become subject to merit retention (i.e., who were already in office when the Article 255 panel was created) will work more quickly but produce lower-quality judgments after becoming subject to merit retention than before. The marginal effect of RETENTION on DURATION, WORD COUNT, and CITATIONS will be negative.

Our second hypothesis is based on Result 3, which predicts that effort is decreasing in the commission's cost of an unfavorable recommendation, c. This cost is higher when a judge has been subject to merit selection (i.e., when a merit committee has already approved the

judge). For the Article 255 panel, changing it's mind about a judge's competence not only risks embarrassing the appointing member state, it also is an admission that the panel was wrong to recommend the judge in the first place. This, in turn, allows rational judges to adjust their efforts downwards without the risk of deselection. Therefore, we expect that judges who were subject to merit selection (SELECTION = 1) will be less productive but produce higher-quality judgments than judges that were appointed before the creation of the panel (SELECTION = 0).

Hypothesis 2. Judges who have been subject to merit selection (i.e., who were appointed after the creation of the Article 255 panel), will work more slowly but produce higher-quality judgments than judges who are only subject to merit retention (i.e., who were appointed before the creation of the panel). The marginal effect of SELECTION on DURATION, WORD COUNT, and CITATIONS will be positive.

Our third hypothesis is based on Result 4, which predicts that effort is decreasing in the expected competence of the judge, $\overline{\omega}$. At the CJEU, we expect that the competence of judges at the CJ will be higher than at the GC, but we cannot compare the two directly because of the many institutional differences. However, according to Result 4, it is more likely that a semi-separating equilibrium exists at the General Court than at the Court of Justice. At the Court of Justice, a pooling equilibrium is more likely to exist, in which the panel is always willing to give a favorable recommendation because the competence of Court of Justice judges is high in expectation. The empirical implication is that we should expect the relationships in Hypotheses 1 and 2 to be weaker at the Court of Justice than at the General Court. In fact, if the expected competence of Court of Justice judges is sufficiently high, the model predicts that these relationships will not hold.

Hypothesis 3. The relationships in Hypotheses 1 and 2 are more likely to hold at the General Court (i.e., where the expected competence of judges is lower) than at the Court of Justice (i.e., where the expected competence of judges is higher).

Estimation Strategy

To test Hypothesis 1, we estimate the marginal effect of RETENTION on three operational-izations of effort: DURATION, WORD COUNT, and CITATIONS. For each dependent variable, we estimate linear models with heteroskedasticity-robust standard errors.[3] For this test, we estimate separate models for each court and restrict our sample to cases in which the judge-rapporteur served on the bench both before and after the creation of the Article 255 panel on 1 March, 2010. The sample for the Court of Justice includes 4,786 cases and the sample for the General Court includes 3,085 cases. Formally, the OLS model for our analysis where i indexes the case is:

$$\mathbf{Y}_i = \beta_0 + \beta_1 \cdot Retention + \delta\mathbf{X} + \psi + \lambda + \epsilon_{ipt} \qquad (1)$$

with $\mathbf{Y}_i$ a vector of the dependent variables (DURATION, WORD COUNT, CITATIONS), $\delta\mathbf{X}_i$ a vector of control variables, ψ_p judge-rapporteur fixed-effects, λ_t year fixed-effects, and ϵ_{ipt} standard errors clustered by judge-rapporteur and year. The judge-rapporteur fixed-effects and year fixed-effects control for unobserved heterogeneity over time and across judge-rapporteurs. This specification only leverages within-judge variation to estimate the marginal effect of RETENTION. In this specification, we compare how each judge performs before and after the creation of the Article 255 panel, which captures the increase in the standards for reappointment associated with the creation of the panel. Including year fixed-effects reduces the risk that our RETENTION dummy captures unobserved heterogeneity over time.

To test Hypothesis 2, we estimate the marginal effect of SELECTED on DURATION, WORD COUNT, and CITATIONS. Again, we estimate linear models with heteroskedasticity-robust standard errors. For this test, we restrict our sample to cases lodged after the creation of the

---

[3] We estimate linear models so we can include fixed-effects without introducing bias due to the incidental parameters problem. We think controlling for unobserved heterogeneity is more important to the inferences we are interested in making than correctly modeling the distribution of the dependent variable. In addition, since our independent variables of interest are dichotomous, the assumption we make about the distribution of the dependent variable is not as critical.

panel. Thus, the sample includes all cases lodged between 1 March, 2010 and 31 December, 2019. Again, we estimate separate models for the Court of Justice and the General Court. The sample for the Court of Justice includes 2,497 cases, and the sample for the General Court includes 2,880 cases. Formally, the OLS model for our analysis where i indexes the case is:

$$\mathbf{Y}_i = \beta_0 + \beta_1 \cdot Selection + \delta\mathbf{X} + \mu + \lambda + \epsilon_{ist} \tag{2}$$

with $\mathbf{Y}_i$ a vector of the dependent variables (DURATION, WORD COUNT, CITATIONS), $\delta\mathbf{X}_i$ a vector of control variables, μ_s member state fixed-effects, λ_t year fixed-effects, and ϵ_{ist} standard errors clustered by member state and year. We include member state fixed-effects and year fixed-effects to control for unobserved heterogeneity over time and across member states. In this specification, we compare judge-rapporteurs appointed after the creation of the Article 255 panel to judge-rapporteurs appointed before its creation in the post-treatment period (i.e., after the creation of the panel), during which all judges are subject to merit retention. We therefore do not include judge-rapporteur fixed effects.

All of our models control for the complexity of the case, which we expect correlates with both our independent and dependent variables thus confounding our estimates of the marginal effects of RETENTION and SELECTION. For the models on the CJ, we also control for whether the CJ requests an AG opinion (AG OPINION), which can prolong cases. We expect complexity to be positively correlated with DURATION, WORD COUNT, and CITATIONS, as complex cases are likely to take longer to decide, require longer judgments, and include more citations. We also expect that complexity will be positively correlated with our independent variables of interest, RETENTION and SELECTION. The complexity of cases has generally increased over time, leading to a positive correlation with RETENTION. In addition, the President of the Court, who decides which cases to assign to which judge-rapporteurs, could disproportionately assign more complex cases to judges who were appointed after the creation

of the Article 255 panel. These judges were subject to merit selection, and may therefore be more competent (if less experienced) than existing judges. This could lead to a positive correlation with SELECTION.

We develop several measures of complexity that capture three major dimensions of the concept: political, policy and legal complexity. Political complexity (or political salience) is captured by the size of the panel that hears the case (PANEL SIZE), as the Court tends to reserve larger panels for more politically salient cases. Policy complexity is approximated by the the number of policy areas related to the case (POLICY AREAS). Finally, legal complexity is measured in three ways: i) A dummy that captures whether the case deals with fundamental principles of law (PRINCIPLES OF LAW). ii) A dummy of whether the case deals with the EU legal order (LEGAL ORDER), and iii) a count of the sources of law that the case deals with. Sources of law include primary law (the EU treaties), secondary law (legislation), and case law. Cases that deal with fundamental principles of law should be more straightforward, because the law is more established. Cases that deal with the legal order of the EU, including the relationship between EU law and national law, tend to be more complex. Cases that deal with multiple sources of law tend to be more complex because they require a greater degree of synthesis. Lastly, we include controls for the type of proceeding before the court PRELIMINARY RULING, ANNULMENTS, INFRINGEMENT, and APPEALS. Note, preliminary rulings and infringement cases are heard only at the CJ and not at the GC.

Findings

Figure 2 illustrates the marginal effects from our models. See Supporting Information Section A for full regression tables. To test Hypothesis 1, the first column shows our estimated marginal effects for RETENTION. To test Hypothesis 2, the second column shows estimated marginal effects for SELECTION. The vertical labels on the right of the plot denote each operationalization of effort (dependent variable; DURATION, WORD COUNT, CITATIONS).
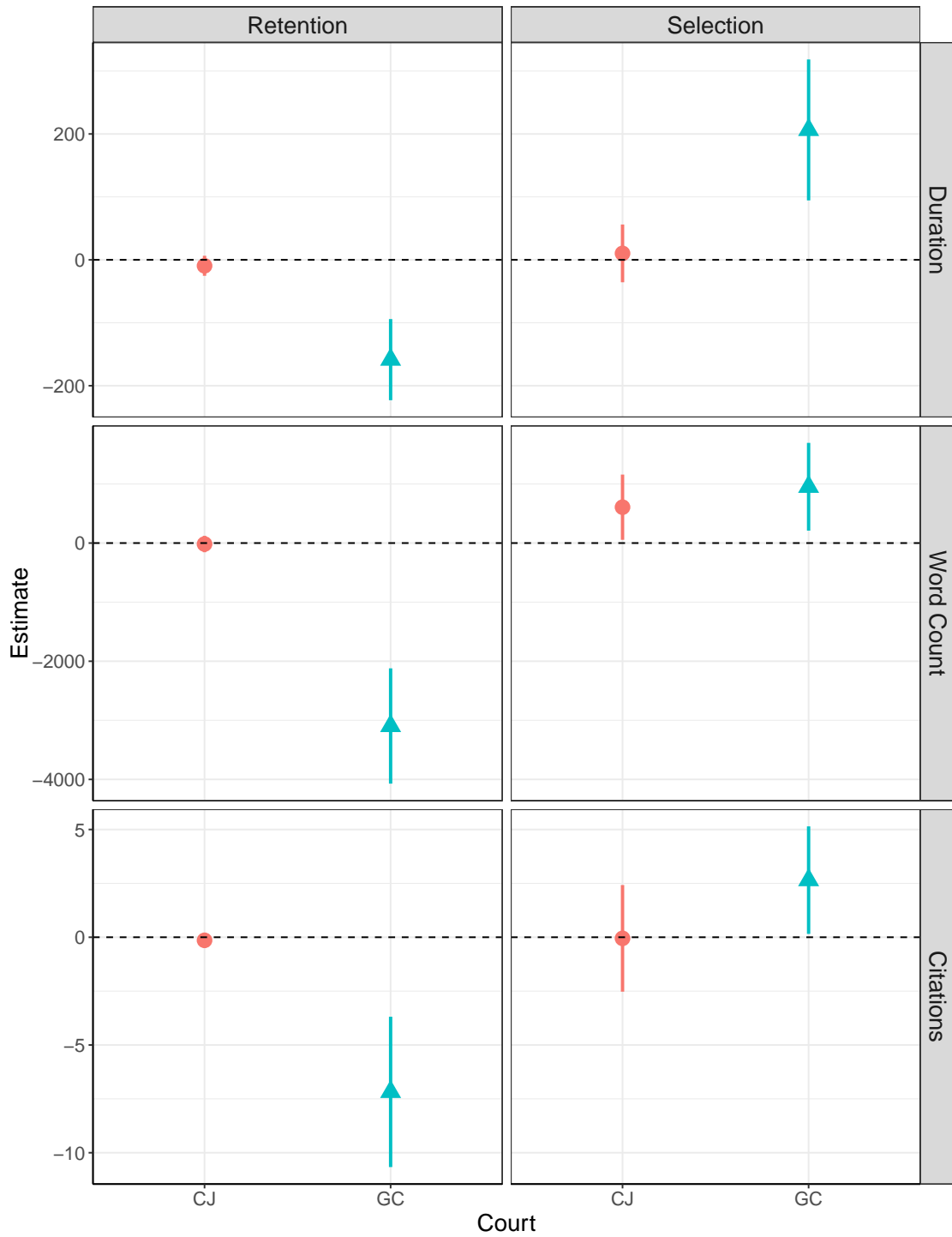
Figure 2. Estimated marginal effects.

Note: This figure shows the estimated marginal effects of RETENTION and SELECTION on our three dependent variables: DURATION, WORD COUNT, and CITATIONS.

To test Hypothesis 3, each panel contrasts the marginal effects for the CJ and the GC; labelled separately on the x-axis.

Looking at the first column, we find that the marginal effect of RETENTION on effort is negative and statistically significant for all three measures at the General Court. By contrast, at the Court of Justice, the estimated marginal effect of RETENTION is not only statistically insignificant for all three measures, it is also substantively negligible. These results are consistent with Hypothesis 1 and Hypothesis 3, predicting that a semi-separating equilibrium is more likely at the GC than the CJ. Specifically, we find that judges that were already in office at the GC when the panel was created, worked more quickly (DURATION), but produced lower-quality judgments (WORD COUNT and CITATIONS) once merit retention was introduced. On average, cases are nearly half a year shorter, with a reduction of 159 days. Additionally, judgments are 3,096 words shorter and contain 7 fewer case law citations. These are substantial effects: The average time to decide a case in the sample is 811 days, the average judgment consists of 10,192 words, typically containing 17 case law citations.

Looking at the second column, we also find that the marginal effect of SELECTION is positive and statistically significant at the GC, but not at the CJ. The one exception is the model for WORD COUNT, for which selection is also positive and statistically significant at the CJ. These results are consistent with Hypothesis 2 and Hypothesis 3. Judges that were appointed after the creation of the panel – and thus were selected by the panel already at the outset – work more slowly and produce higher-quality judgments than judges who are only subject to merit retention (i.e, who were appointed before the creation of the panel). On average, their cases last 206 days longer, their judgments are 953 words longer, and their judgments have 3 more case law citations. Again, these effects are substantively large.

In sum, we find strong empirical support for the testable predictions of our formal model. Consistent with Hypothesis 1, merit retention without initial selection causes GC judges to expend more effort to satisfy the criteria of the Article 255 panel, leading to shorter, lower-

quality judgments. Consistent with Hypothesis 2, merit selection leads GC judges to expend less effort, leading to longer, higher-quality judgments. Lastly, consistent with Hypothesis 3, the Article 255 panel only influences the behavior of judges at the GC, where judges are expected to be relatively less competent than their colleagues at the CJ.

Conclusion

How do merit commissions affect judicial behavior? Our findings suggest that merit commissions motivate judges in counterintuitive ways. Retention incentivizes judges to expend more effort on meeting the criteria of a merit commission, but merit selection incentivizes them to expend less. Merit commissions can also have unintended consequences, depending on the criteria they establish for reappointment. By focusing on easily measurable criteria, such as productivity, the Article 255 panel has reduced the quality of some General Court judgments, and increased the quality of others. Our findings make a number of contributions to the literature.

First, we build upon an extensive scholarship on how institutional selection and retention mechanisms affect judicial behavior (e.g., Canes-Wrone, Clark and Kelly 2014; Tiede 2023). In particular, we compare the functions of merit commissions in both common law and civil law contexts. As Garoupa and Ginsburg (2009, 119) note, "Academic work on judicial councils has been so far quite limited. There are very few empirical studies and there has been no [...] statistical analysis to date". By leveraging the 255 Panel's timing and institutional configuration, we are able to directly compare the consequences of a merit commission focusing primarily on selection to one focusing on both selection and retention. One implication of our findings is that merit commissions that select candidates may not be suitable for evaluating the retention of the same candidates, as their evaluations may be biased in favor of incumbents and against acknowledging that their previous selection was a mistake.

Second, we contribute to the scholarship on judicial selection and independence in the context of international courts (e.g., Dunoff and Pollack 2017; Voeten 2007; Stiansen 2022). Although the merit commission at the CJEU is relatively unique among international courts, our findings provide insights on the tradeoffs other international organizations may face by implementing such a monitoring institution for appointments to their court. Importantly, while no EU member state has yet continued with the appointment of a judge following a negative recommendation by the 255 panel, if some member states were to ignore the panel's decision-making together, a future theory can build upon our insights and examine whether the theoretical expectations and empirical implications we derive would still exist.

Lastly, following the recommendation of Staton and Moore (2011), this article draws upon the American, comparative, and international courts scholarship. By integrating these sub-disciplinary insights, scholars can converse more effectively about the institutional mechanisms that affect judicial decision-making and produce theoretical innovations that can be empirically testable across domestic and international contexts. Our leveraging of the scholarship on merit commissions is one example of this analytical tractability across the subfields. Future research can similarly benefit by borrowing and building upon insights generated through the analysis of either international or domestic courts.

References

Avril, Lola. 2022. The Genesis of the Institution within the Institution: Studying the Mobilization for the Creation of the Court of First Instance. In *Researching the European Court of Justice: Methodological Shifts and Law's Embeddedness*, ed. Antoine Vauchez, Fernanda Nicola and Mikael Rask Madsen. Studies on International Courts and Tribunals Cambridge University Press pp. 237–260.

Bielen, Samantha, Ludo Peeters, Wim Marneffe and Lode Vereeck. 2018. "Backlogs and litigation rates: Testing congestion equilibrium across European judiciaries." *International Review of Law and Economics* 53:9–22.

Bonica, Adam and Maya Sen. 2017. "The politics of selecting the bench from the bar: The legal profession and partisan incentives to introduce ideology into judicial selection." *Journal of Law and Economics* 60(4):559–595. Publisher: University of Chicago Press.

Brekke, Stein Arne, Joshua C Fjelstul, Silje Synnøve Lyder Hermansen and Daniel Naurin. 2023. "The CJEU Database Platform: Decisions and Decision-Makers." *The Journal of Law and Courts* .

Bueno de Mesquita, Ethan and Matthew Stephenson. 2007. "Regulatory Quality Under Imperfect Oversight." *American Political Science Review* 101(3):605–620.

Canes-Wrone, Brandice, Tom S Clark and Jason P Kelly. 2014. "Judicial Selection and Death Penalty Decisions." *American Political Science Review* 108(1):23–39.

Canes-Wrone, Brandice, Tom S Clark and Jee Kwang Park. 2012. "Judicial independence and retention elections." *Journal of Law, Economics, and Organization* 28(2):211–234.

Cheruvu, Sivaram. 2019. "How do institutional constraints affect judicial decision-making? The European Court of Justice's French language mandate." *European Union Politics* 20(4):562–583.

Choi, Stephen J, G Mitu Gulati and Eric A Posner. 2010. "Professionals or Politicians: The Uncertain Empirical Case for an Elected Rather than Appointed Judiciary." *The Journal of Law Economics, & Organization* 26(2):290–336.

De Waele, Henri. 2021. "Re-Appraising Success and Failure in the Life of the European Court of Justice." *Cambridge Yearbook of European Legal Studies* 23:54–72.

Dehousse, Franklin. 2011. "The Reform of the EU Courts: The Need of a Management Approach." *Egmont Paper* (53):1–38.

Dehousse, Franklin. 2016. "The Lessons for Judicial Management.".

Devaney, James. 2019. "An Independent Panel for the Scrutiny of Investment Arbitrators: an Idea Whose Time Has Come?" *The Law & Practice of International Courts and Tribunals* 18:369–388.

Dumbrovsky, Tomas, Bilyana Petkova and Marijn Van Der Sluis. 2014. "Judicial Appointments: The Article 255 TFEU Advisory Panel and Selection Procedures in the Member States." *Common Market Law Review* 51:455–482.

Dunoff, Jeffrey L and Mark A Pollack. 2017. "The Judicial Trilemma." *American Journal of International Law* 111(2):225–276. Publisher: Cambridge University Press.

Epstein, Lee, William M Landes and Richard A Posner. 2011. "Why (and When) Judges Dissent: A theoretical and empirical analysis." *Journal of Legal Analysis* 3(1):101–138.

Fjelstul, Joshua. 2022. "How the Chamber System at the CJEU Undermines the Consistency of the Court's Application of EU Law." *Journal of Law and Courts* 10(2).

Garoupa, Nuno and Tom Ginsburg. 2009. "Guarding the Guardians: Judicial Councils and Judicial Independence." *American Journal of Comparative Law* 57(1):103–134. Publisher: American Society of Comparative Law.

Goelzhauser, Greg. 2018. "Does Merit Selection Work?: Evidence from Commission and Gubernatorial Choices." *Journal of Law and Courts* 6(1):155–187.

Gordon, Sanford C and Gregory A Huber. 2007. "The Effect of Electoral Competitiveness on Incumbent Behavior." *Quarterly Journal of Political Science* 2(2):107–138. Publisher: Now Publishers, Inc.

Hermansen, Silje Synnøve Lyder. 2020. "Building legitimacy: strategic case allocations in the Court of Justice of the European Union." *Journal of European Public Policy* 27(8):1215–1235.

Hermansen, Silje Synnøve Lyder and Daniel Naurin. 2019. Holding Judges To Account. Policy, Performance and Impartiality. In *Conference Paper*.

Kritzer, Herbert M. 2016. "Impact of Judicial Elections on Judicial Decisions." *Annual Review of Law and Social Science* 12:353–371. Publisher: Annual Reviews.

Larsson, Olof, Daniel Naurin, Mattias Derlén and Johan Lindholm. 2017. "Speaking Law to Power: The Strategic Use of Precedent of the Court of Justice of the European Union." *Comparative Political Studies* 50(7):879–907.

Lupu, Yonatan. 2013. "International Judicial Legitimacy: Lessons from National Courts." *Theoretical Inquiries in Law* 14(2):437–454.

Owens, Ryan J, Alexander Tahk, Patrick C Wohlfarth and Amanda C Bryan. 2015. "Nominating Commissions, Judicial Retention, and Forward-Looking Behavior on State Supreme Courts: An Empirical Examination of Selection and Retention Methods." *State Politics & Policy Quarterly* 15(2):211–238.

Pérez, Aida Torres. 2015. Can Judicial Selection Secure Judicial Independence?: Constraining State Governments in Selecting International Judges. In *Selecting Europe's Judges: A*

Critical Review of the Appointment Procedures to the European Courts, ed. Michal Bobek. Oxford University Press pp. 181–201.

Ramseyer, J Mark and Eric B Rasmusen. 2003. *Measuring Judicial Independence: The Political Economy of Judging in Japan.* Chicago: University of Chicago Press.

Sauvé, Jean-Marc. 2015. Selecting the European Union's Judges: The Practice of the Article 255 Panel. In *Selecting Europe's Judges*, ed. Michal Bobek. Oxford University Press.

Staton, Jeffrey K and Will H Moore. 2011. "Judicial Power in Domestic and International Politics." *International Organization* 65(3):553–587.

Stiansen, Øyvind. 2022. "(Non)renewable Terms and Judicial Independence in the European Court of Human Rights." *Journal of Politics* 84(2):992–1006.

Tiede, Lydia Brashear. 2023. Selecting Judges. In *Oxford Handbook of Comparative Judicial Behaviour*, ed. Lee Epstein, Gunnar Grendstad, Urška Šadl and Keren Weinshall. Oxford: Oxford University Press.

Tobias, Carl W. 2008. "Reconsidering Virginia Judicial Selection." *University of Richmond Law Review* 43:37–50.

Voeten, Erik. 2007. "The Politics of International Judicial Appointments: Evidence from the European Court of Human Rights." *International Organization* 61(4):669–701.

von Bogdandy, Armin and Christoph Krenn. 2015. On the Democratic Legitimacy of Europe's Judges: A Principled and Comparative Reconstruction of the Selection Procedures. In *Selecting Europe's Judges: A Critical Review of the Appointment Procedures to the European Courts*, ed. Michal Bobek. Oxford University Press p. 0.

Supplemental Appendix for "How do merit commissions affect judicial behavior? Evidence from the Court of Justice of the European Union"

Appendices

Table of Contents

A   Regression tables

Tables A1 (Duration), A2 (Word Count), and A3 (Citations) present our full regression results.

Table A1. Duration as the Dependent Variable

Dependent Variable:		Duration		
Court:	CJ	GC	CJ	GC
Model:	(1)	(2)	(3)	(4)
Variables				
Retention	-9.470	-158.6***		
	(8.109)	(32.88)		
Selection			10.25	206.4***
			(23.41)	(57.12)
Panel Size	13.82***	22.23	5.393***	105.1***
	(2.691)	(23.50)	(1.232)	(24.83)
Policy Areas	62.54***	38.38*	36.50**	26.94
	(10.05)	(20.20)	(12.61)	(19.70)
Principles of Law	-0.4640	3.808	3.941	60.55*
	(14.84)	(39.04)	(9.212)	(27.06)
Legal Order	22.32	246.2***	10.47	137.9***
	(17.11)	(28.75)	(10.26)	(10.54)
Sources of Law	12.41	208.6***	20.33***	111.7***
	(10.94)	(29.12)	(5.601)	(30.20)
AG Opinion	128.2***		131.9***	
	(11.01)		(10.90)	
Preliminary Ruling	17.24		35.35	
	(36.26)		(51.00)	
Annulments	193.9***	-19.64	199.6***	-95.08*
	(45.89)	(45.99)	(56.53)	(42.06)
Infringement	57.37		118.1*	
	(39.35)		(52.95)	
Appeals	107.2**	-188.1***	129.7**	-182.1***
	(41.91)	(48.18)	(52.54)	(39.82)
Fixed-effects				
Judge Rapporteur	Yes	Yes		
Case Year	Yes	Yes	Yes	Yes
Member State			Yes	Yes
Fit statistics				
SE Clustering	JR + Year	JR + Year	MS + Year	MS + Year
R^2	0.41510	0.62294	0.37053	0.40756
Observations	4,786	2,497	3,085	2,879

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: JR = Judge Rapporteur, MS = Member State

Table A2. Word Count as the Dependent Variable

Dependent Variable:	Word Count			
Court:	CJ	GC	CJ	GC
Model:	(1)	(2)	(3)	(4)
Variables				
Retention	-16.75	-3,095.9***		
	(70.61)	(497.7)		
Selection			606.9*	952.6**
			(281.6)	(378.8)
Panel Size	214.8***	2,533.3*	139.6**	3,694.5***
	(28.92)	(1,289.5)	(50.81)	(1,076.1)
Policy Areas	1,393.5***	2,578.7***	1,134.0***	1,786.3**
	(91.32)	(321.7)	(147.2)	(595.3)
Principles of Law	1,263.6***	2,495.7*	1,115.5***	3,398.1***
	(297.2)	(1,334.7)	(326.9)	(564.0)
Legal Order	692.6***	6,211.4***	650.1***	3,244.3***
	(141.3)	(1,508.6)	(146.5)	(298.5)
Sources of Law	235.7	3,893.1***	734.5***	2,604.0***
	(143.9)	(446.8)	(90.97)	(465.5)
AG Opinion	716.6***		573.3**	
	(116.1)		(247.7)	
Preliminary Ruling	-873.6*		-1,357.5***	
	(453.0)		(170.5)	
Annulments	99.04	-1,297.4	-360.4	-1,690.8
	(625.5)	(1,240.9)	(414.5)	(1,260.8)
Infringement	-989.9**		-583.0*	
	(425.1)		(289.1)	
Appeals	1,872.7***	-396.8	796.5*	-233.3
	(585.3)	(1,342.6)	(425.5)	(1,105.4)
Fixed-effects				
Judge Rapporteur	Yes	Yes		
Case Year	Yes	Yes	Yes	Yes
Member State			Yes	Yes
Fit statistics				
SE Clustering	JR + Year	JR + Year	MS + Year	MS + Year
R^2	0.34988	0.30983	0.29350	0.27246
Observations	4,786	2,497	3,085	2,879

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: JR = Judge Rapporteur, MS = Member State

Table A3. Citations as a Dependent Variable

Dependent Variable:	Citations			
Court:	CJ	GC	CJ	GC
Model:	(1)	(2)	(3)	(4)
Variables				
Retention	-0.1432	-7.177***		
	(0.1482)	(1.780)		
Selection			-0.0514	2.651*
			(1.262)	(1.274)
Panel Size	0.2001***	1.240	0.2131**	4.421**
	(0.0451)	(1.332)	(0.0732)	(1.672)
Policy Areas	2.312***	7.116***	1.472***	4.735***
	(0.2170)	(0.7916)	(0.2212)	(0.8105)
Principles of Law	2.184***	4.707	2.312***	3.743***
	(0.4320)	(2.769)	(0.6395)	(0.5423)
Legal Order	1.934***	6.204***	1.463***	1.438
	(0.3601)	(2.099)	(0.2432)	(1.214)
Sources of Law	2.465***	7.821***	2.881***	5.715***
	(0.1693)	(0.7682)	(0.2057)	(0.3859)
AG Opinion	1.103***		0.6798***	
	(0.2493)		(0.1903)	
Preliminary Ruling	-1.765		-1.636**	
	(1.942)		(0.5979)	
Annulments	-3.184	-0.6867	-4.314***	-3.282*
	(2.143)	(2.009)	(1.009)	(1.734)
Infringement	-1.501		-1.326*	
	(1.872)		(0.5939)	
Appeals	1.080	1.333	1.446**	-3.745
	(1.895)	(2.812)	(0.5679)	(2.974)
Fixed-effects				
Judge Rapporteur	Yes	Yes		
Case Year	Yes	Yes	Yes	Yes
Member State			Yes	Yes
Fit statistics				
SE Clustering	JR + Year	JR + Year	MS + Year	MS + Year
R^2	0.33444	0.34226	0.27293	0.22266
Observations	4,786	2,497	3,085	2,879

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: JR = Judge Rapporteur, MS = Member State

B Model presentation

- Players:

    - J is a judge whose reappointment depends on a favorable recommendation from a merit commission.

    - C is a merit commission.

- Order of play:

    - Nature chooses the type of the judge, $\omega \sim \mathcal{U}(\underline{\omega}, \overline{\omega})$, which is private information to the judge.

    - The judge chooses a level of effort, $e > 0$, to expend on satisfying the criteria established by the merit commission.

    - The merit commission observes the effort expended by the judge, e, and chooses a recommendation, r. The commission can issue a favorable recommendation, F, or a negative recommendation, U.

    - The game ends and the players receive their payoffs.

- Strategies:

    - A strategy for the judge, $\sigma_{J|\omega}(\sigma_C)$, is a function that maps a level of effort, $e > 0$, conditional on their type, ω, to the action of the merit commission, $r \in \{F, U\}$.

    - A strategy for the merit commission, $\sigma_C(\sigma_{J|\omega})$, is a function that maps a recommendation $r \in \{F, U\}$, to the action of the judge, e.

- Assumptions about preferences:

    - Judges are career-motivated and want to be reappointed. By law or norm, reappointment requires a favorable recommendation from the merit commission.

    - Meeting the reappointment criteria of a merit commission requires effort, which is costly for judges.

    - The marginal cost of exerting additional effort is lower for high-competence judges than for low-competence judges.

    - Merit commissions want to give favorable recommendations to sufficiently competent judges and unfavorable recommendations to insufficiently competent judges.

- It is politically costly for a merit commission to reject a judge because there could be backlash from the appointing government.

- Utility functions:

  - The judge receives a benefit, b, if the merit commission chooses to give a favorable recommendation.

  - The judge pays a cost proportional to the level of effort they choose to expend, $-\frac{e}{\omega}$. This cost is smaller for higher-quality types and larger for low-quality types. This captures the intuition that high-quality types can expend more effort at a lower cost than low-quality types.

  $$u_{J|\omega}(F \mid e) = b - \frac{e}{\omega}$$
  $$u_{J|\omega}(U \mid e) = -\frac{e}{\omega}$$

  - If the merit commission gives a favorable recommendation, the judge gets reappointed, and the commission receives a cost or benefit, $\omega - \omega_i$, depending on the type of the judge. The commission does not know the type of the judge, but has a belief about it. The commission has a minimal standard for judge quality, ω_i. If the type of the judge exceeds that standard, $\omega > \omega_i$, then the commission receives a benefit for giving a favorable recommendation. However, if the type of the judge falls short of that standard, $\omega < \omega_i$, the commission pays a cost.

  - If the merit commission gives a negative recommendation, it pays a cost, $-c$. This cost captures the political cost of calling out the government that appointed the judge. This cost is higher if the merit commission screened the judge, as giving a negative recommendation to a judge the commission has already approved is an implicit admission that the commission got it wrong.

  $$u_C(F \mid \omega) = \omega - \omega_i$$
  $$u_C(U \mid \omega) = -c$$

- Endogenous choice variables:

  - $e > 0$ is the level of observable effort that the judge chooses to expend in order to satisfy the reappointment criteria of the merit commission.

  - $r \in \{F, U\}$ is the recommendation of the merit commission. The commission can give a favorable recommendation, F, or an unfavorable recommendation, U.

7

- Exogenous parameters:

  - $\omega \sim \mathcal{U}(\underline{\omega}, \overline{\omega})$ is the type of the judge, where $\mathcal{U}$ is the continuous uniform distribution, $\underline{\omega}$ is the lower bound of the distribution, and $\overline{\omega}$ is the lower bound of the distribution.

  - $\omega_i \in (\underline{\omega}, \overline{\omega})$ is the minimum judge type that is acceptable to the merit commission. Looking at $u_C(F \mid \omega) = \omega - \omega_i$, when $\omega < \omega_i$, giving a favorable recommendation is costly to the merit commission. When $\omega > \omega_i$, giving a favorable recommendation is beneficial.

  - $b > 0$ is the benefit to the judge of receiving a favorable recommendation from the merit commission.

  - $c > 0$ is the political cost to the merit commission for giving an unfavorable recommendation.

- Information structure:

  - The type of the judge, ω, is private information. The merit commission knows the distribution, $\omega \sim \mathcal{U}(\underline{\omega}, \overline{\omega})$. All other parameters are common knowledge.

Equilibrium Analysis

- Candidate equilibrium:

  - We posit that this game has a unique semi-separating perfect Bayesian equilibrium (PBE) with two on-path signals. We can write this equilibrium as a tuple, $\{\sigma_{J|\omega}(\sigma_C), \sigma_C(\sigma_{J|\omega}), \mu_C(\omega \mid e)\}$, where $\sigma_{J|\omega}(\sigma_C)$ is the strategy profile of the judge, $\sigma_C(\sigma_{J|\omega})$ is the strategy profile of the merit commission, and $\mu_C(\omega \mid x)$ are the beliefs of the merit commission.

  - Formally, the strategies and beliefs of the players in the candidate equilibrium are as follows:

$$
\sigma_{J|\omega}(\sigma_C) = \begin{cases} e = \overline{e}^* & \text{if } \omega \geq \omega^* \\ e = \underline{e}^* = 0 & \text{if } \omega < \omega^* \end{cases}
$$

$$
\sigma_C(\sigma_{J|\omega}) = \begin{cases} r = F & \text{if } e \geq \overline{e}^* \\ r = U & \text{if } e < \overline{e}^* \end{cases}
$$

$$\mu_C(\omega \mid e) = \begin{cases} \omega \sim \mathcal{U}(\omega^*, \overline{\omega}) & \text{if } e = \overline{e}^* \\ \omega \sim \mathcal{U}(\underline{\omega}, \omega^*) & \text{if } e = \underline{e}^* = 0 \end{cases}$$

- Sufficiently competent judges, $\omega \geq \omega^*$, choose a high level of effort, $\overline{e}^*$, and insufficiently competent judges, $\omega < \omega^*$, choose a low level of effort, $\underline{e}^* = 0$.

- The prior belief of the merit commission are that the type of the judge, ω, is uniformly distributed between a lower bound, $\underline{\omega}$, and an upper bound, $\overline{\omega}$: $\omega \sim \mathcal{U}(\underline{\omega}, \overline{\omega})$.

- Upon observing the strong signal $e = \overline{e}^*$, the merit commission believes that the type of the judge, ω, is uniformly distributed between a cut-point, ω^*, and the upper bound, $\overline{\omega}$: $\omega \sim \mathcal{U}(\omega^*, \overline{\omega})$. Based on this belief, the merit commission makes a favorable recommendation, F.

- Upon observing a the weak signal, $e = \underline{e}^* = 0$, the merit commission believes that the type of the judge, ω, is uniformly distributed between the lower bound, $\underline{\omega}$, and a cut-point, ω^*: $\omega \sim \mathcal{U}(\omega^*, \overline{\omega})$. Based on this belief, the merit commission makes an unfavorable recommendation, U.

- Proof of equilibrium existence:

  - Step 1: Are the on-path actions in the candidate equilibrium sequentially rational for the commission?

    - Condition 1: The commission has to prefer to give a favorable recommendation when it observes the strong signal, $\overline{e}^*$.
    - Beliefs: When the commission observes the strong signal, $\overline{e}^*$, it believes that the type of the judge, ω, is uniformly distributed between the equilibrium cut-point, ω^*, and the upper bound, $\overline{\omega}$: $\omega \sim \mathcal{U}(\omega^*, \overline{\omega})$. The commission then gives a favorable recommendation, F.

$$\mathbb{E}\big[u_C(F \mid \overline{e}^*)\big] \geq \mathbb{E}\big[u_C(U \mid \overline{e}^*)\big]$$

$$\int_{\omega^*}^{\overline{\omega}} u_C(F \mid \overline{e}^*) f(\omega) \, \mathrm{d}\omega \geq \int_{\omega^*}^{\overline{\omega}} u_C(U \mid \overline{e}^*) f(\omega) \, \mathrm{d}\omega$$

$$\int_{\omega^*}^{\overline{\omega}} (\omega - \omega_i) \left(\frac{1}{\overline{\omega} - \underline{\omega}}\right) \mathrm{d}\omega \geq \int_{\omega^*}^{\overline{\omega}} -c \left(\frac{1}{\overline{\omega} - \underline{\omega}}\right) \mathrm{d}\omega$$

$$\int_{\omega^*}^{\overline{\omega}} \frac{\omega - \omega_i}{\overline{\omega} - \underline{\omega}} \, \mathrm{d}\omega \geq \int_{\omega^*}^{\overline{\omega}} \frac{-c}{\overline{\omega} - \underline{\omega}} \, \mathrm{d}\omega$$

$$\frac{(\overline{\omega} - \omega^*)(\overline{\omega} + \omega^* - 2\omega_i)}{2(\overline{\omega} - \underline{\omega})} \geq \frac{-c(\overline{\omega} - \omega^*)}{\overline{\omega} - \underline{\omega}}$$

$$\overline{\omega} + \omega^* - 2\omega_i \geq -2c$$

$$c \geq \frac{2\omega_i - \omega^* - \overline{\omega}}{2}$$

- Condition 2: The commission has to prefer to give an unfavorable recommendation when it observes the weak signal, $\underline{e}^*$.

- Beliefs: When the commission observes the weak signal, $\underline{e}^*$, it believes that the type of the judge, ω, is uniformly distributed between the lower bound, $\underline{\omega}$, and the equilibrium cut-point, ω^*: $\omega \sim \mathcal{U}(\underline{\omega}, \omega^*)$. The commission then gives an unfavorable recommendation, U.

$$\mathbb{E}\big[u_C(U \mid \underline{e}^*)\big] \geq \mathbb{E}\big[u_C(F \mid \underline{e}^*)\big]$$

$$\int_{\underline{\omega}}^{\omega^*} u_C(U \mid \underline{e}^*) f(\omega) \, \mathrm{d}\omega \geq \int_{\underline{\omega}}^{\omega^*} u_C(F \mid \underline{e}^*) f(\omega) \, \mathrm{d}\omega$$

$$\int_{\underline{\omega}}^{\omega^*} -c\left(\frac{1}{\overline{\omega} - \underline{\omega}}\right) \mathrm{d}\omega \geq \int_{\underline{\omega}}^{\omega^*} (\omega - \omega_i)\left(\frac{1}{\overline{\omega} - \underline{\omega}}\right) \mathrm{d}\omega$$

$$\int_{\underline{\omega}}^{\omega^*} \frac{-c}{\overline{\omega} - \underline{\omega}} \, \mathrm{d}\omega \geq \int_{\underline{\omega}}^{\omega^*} \frac{\omega - \omega_i}{\overline{\omega} - \underline{\omega}} \, \mathrm{d}\omega$$

$$\frac{-c(\omega^* - \underline{\omega})}{\overline{\omega} - \underline{\omega}} \geq \frac{(\omega^* - \underline{\omega})(\omega^* + \underline{\omega} - 2\omega_i)}{2(\overline{\omega} - \underline{\omega})}$$

$$-2c \geq \omega^* + \underline{\omega} - 2\omega_i$$

$$c \leq \frac{2\omega_i - \omega^* - \underline{\omega}}{2}$$

- Intuition: Upon seeing the strong signal, the merit commission is willing to give a favorable ruling, despite its uncertainty about the type of the judge, as long as the cost of an unfavorable ruling is sufficiently high. Upon seeing the weak signal, the merit commission is willing to gave an unfavorable ruling as long as the cost of doing so is sufficiently low.

- Are the conditions compatible?

$$\frac{2\omega_i - \omega^* - \overline{\omega}}{2} \leq \frac{2\omega_i - \omega^* - \underline{\omega}}{2}$$

$$-\overline{\omega} \leq -\underline{\omega}$$

$$\overline{\omega} \geq \underline{\omega}$$

- The conditions are always compatible.
- Step 2: What off-path beliefs are required to make the off-path actions in the candidate equilibrium sequentially rational for the merit commission?

  - Case 1: What off-path beliefs keep the merit commission playing U given an unexpected signal $0 < e' < \overline{e}^*$?

$$\mathbb{E}\big[u_C(U \mid e' < \overline{e}^*)\big] \geq \mathbb{E}\big[u_C(F \mid e' < \overline{e}^*)\big]$$

$$\int_{\underline{\omega}'}^{\overline{\omega}'} u_C(U \mid e') f(\omega)\,\mathrm{d}\omega \geq \int_{\underline{\omega}'}^{\overline{\omega}'} u_C(F \mid e') f(\omega)\,\mathrm{d}\omega$$

$$\int_{\underline{\omega}'}^{\overline{\omega}'} -c\left(\frac{1}{\overline{\omega} - \underline{\omega}}\right)\mathrm{d}\omega \geq \int_{\underline{\omega}'}^{\overline{\omega}'} (\omega - \omega_i)\left(\frac{1}{\overline{\omega} - \underline{\omega}}\right)\mathrm{d}\omega$$

$$\int_{\underline{\omega}'}^{\overline{\omega}'} \frac{-c}{\overline{\omega} - \underline{\omega}}\,\mathrm{d}\omega \geq \int_{\underline{\omega}'}^{\overline{\omega}'} \frac{\omega - \omega_i}{\overline{\omega} - \underline{\omega}}\,\mathrm{d}\omega$$

$$\frac{-c(\overline{\omega}' - \underline{\omega}')}{\overline{\omega} - \underline{\omega}} \geq \frac{(\overline{\omega}' - \underline{\omega}')(\overline{\omega}' + \underline{\omega}' - 2\omega_i)}{2(\overline{\omega} - \underline{\omega})}$$

$$-2c \geq \overline{\omega}' + \underline{\omega}' - 2\omega_i$$

$$c \leq \frac{2\omega_i - \overline{\omega}' - \underline{\omega}'}{2}$$

  - Case 2: What off-path beliefs keep the merit commission playing F given an unexpected signal $e'' > \overline{e}^*$?

$$\mathbb{E}\big[u_C(F \mid e'' > \overline{e}^*)\big] \geq \mathbb{E}\big[u_C(U \mid e'' > \overline{e}^*)\big]$$

$$\int_{\underline{\omega}''}^{\overline{\omega}''} u_C(F \mid e'') f(\omega)\,\mathrm{d}\omega \geq \int_{\underline{\omega}''}^{\overline{\omega}''} u_C(U \mid e'') f(\omega)\,\mathrm{d}\omega$$

11

$$\int_{\underline{\omega}''}^{\overline{\omega}''} (\omega - \omega_i) \left(\frac{1}{\overline{\omega} - \underline{\omega}} \right) \mathrm{d}\omega \geq \int_{\underline{\omega}''}^{\overline{\omega}''} -c \left(\frac{1}{\overline{\omega} - \underline{\omega}} \right) \mathrm{d}\omega$$

$$\int_{\underline{\omega}''}^{\overline{\omega}''} \frac{\omega - \omega_i}{\overline{\omega} - \underline{\omega}} \mathrm{d}\omega \geq \int_{\underline{\omega}''}^{\overline{\omega}''} \frac{-c}{\overline{\omega} - \underline{\omega}} \mathrm{d}\omega$$

$$\frac{(\overline{\omega}'' - \underline{\omega}'')(\overline{\omega}'' + \underline{\omega}'' - 2\omega_i)}{2(\overline{\omega} - \underline{\omega})} \geq \frac{-c(\overline{\omega}'' - \underline{\omega}'')}{\overline{\omega} - \underline{\omega}}$$

$$\overline{\omega}'' + \underline{\omega}'' - 2\omega_i \geq -2c$$

$$c \geq \frac{2\omega_i - \underline{\omega}'' - \overline{\omega}''}{2}$$

- These two conditions create an upper and lower bound for c. When is this range non-empty?

$$\frac{2\omega_i - \underline{\omega}'' - \overline{\omega}''}{2} < \frac{2\omega_i - \underline{\omega}' - \overline{\omega}'}{2}$$

$$-\underline{\omega}'' - \overline{\omega}'' < -\underline{\omega}' - \overline{\omega}'$$

$$\underline{\omega}'' + \overline{\omega}'' > \underline{\omega}' + \overline{\omega}'$$

- We can interpret this condition as follows. The merit commission has to believe that the type of the judge is higher in expectation given an unexpected signal $e > \overline{e}^*$ than given an unexpected signal $0 < e < \overline{e}^*$. This makes the commission willing to play F given $e > \overline{e}^*$ and U given $0 < e < \overline{e}^*$. In other words, the commission has to believe high unexpected signals are more likely to come from high types and lower unexpected signals are more likely to come from low types. Since the marginal cost of a higher signal is lower for higher types, this is a plausible belief.

- Step 3: What is the cut-point on the type of the judge in equilibrium, ω^*, and what are the on-path high and low signals chosen by the judge in equilibrium, $\underline{e}^*$ and $\overline{e}^*$?

  - The cut-point on $\overline{e}^*$ is the level effort required to induce the merit commission to choose F over U.

- The utility of the judge, regardless of type, ω, is strictly decreasing in the level of effort, e:

$$\frac{\partial u_{J|\omega}(e)}{\partial e} < 0$$

- Therefore, the weak signal is $\underline{e}^* = 0$. The judge only ever has an incentive to deviate upward to induce the merit commission to choose F over U, in which case the judge receives b.

- The cut-point on ω^* is the type that is indifferent between sending the weak signal, $\underline{e}^*$, and the strong signal, $\overline{e}^*$.

$$u_{J|\omega^*}(\underline{e}^*) = u_{J|\omega^*}(\overline{e}^*)$$

$$-\frac{\underline{e}^*}{\omega^*} = b - \frac{\overline{e}^*}{\omega^*}$$

$$0 = b - \frac{\overline{e}^*}{\omega^*}$$

$$\frac{\overline{e}^*}{\omega^*} = b$$

$$\omega^* = \frac{\overline{e}^*}{b}$$

- To get a closed-form solution for ω^*, we need to solve for $\overline{e}^*$ and substitute.

- Since the utility of the judge is strictly decreasing in e, the strong signal, $\overline{e}^*$, is the minimum value of e that makes the merit commission choose F, given the commission's beliefs upon observing that signal, $\omega \sim \mathcal{U}(\omega^*, \overline{\omega})$. This is the value of e that makes the commission indifferent between F and U.

$$\mathbb{E}\big[u_C(F \mid \overline{e}^*)\big] = \mathbb{E}\big[u_C(U \mid \overline{e}^*)\big]$$

$$\int_{\omega^*}^{\overline{\omega}} u_C(F \mid \overline{e}^*) f(\omega) \, \mathrm{d}\omega = \int_{\omega^*}^{\overline{\omega}} u_C(U \mid \overline{e}^*) f(\omega) \, \mathrm{d}\omega$$

$$\int_{\omega^*}^{\overline{\omega}} (\omega - \omega_i) \left(\frac{1}{\overline{\omega} - \underline{\omega}}\right) \mathrm{d}\omega = \int_{\omega^*}^{\overline{\omega}} -c \left(\frac{1}{\overline{\omega} - \underline{\omega}}\right) \mathrm{d}\omega$$

$$\int_{\omega^*}^{\overline{\omega}} \frac{\omega - \omega_i}{\overline{\omega} - \underline{\omega}} \, \mathrm{d}\omega = \int_{\omega^*}^{\overline{\omega}} \frac{-c}{\overline{\omega} - \underline{\omega}} \, \mathrm{d}\omega$$

13

$$\frac{(\overline{\omega} - \omega^*)(\overline{\omega} + \omega^* - 2\omega_i)}{2(\overline{\omega} - \underline{\omega})} = \frac{-c(\overline{\omega} - \omega^*)}{\overline{\omega} - \underline{\omega}}$$

$$\overline{\omega} + \omega^* - 2\omega_i = -2c$$

$$\omega^* = 2\omega_i - 2c - \overline{\omega}$$

$$\frac{\overline{e}^*}{b} = 2\omega_i - 2c - \overline{\omega}$$

$$\overline{e}^* = b(2\omega_i - 2c - \overline{\omega})$$

- Now we have closed-form interior solutions for ω^*, $\underline{e}^*$, and $\overline{e}^*$:

$$\omega^* = 2\omega_i - 2c - \overline{\omega}$$

$$\underline{e}^* = 0$$

$$\overline{e}^* = b(2\omega_i - 2c - \overline{\omega})$$

- There is also a boundary solution where $\omega^* = \overline{\omega}$ and $\overline{e}^* = b\overline{\omega}$.
- By assumption, it must be that $\underline{\omega} < \omega^* < \overline{\omega}$ and $e^* > 0$. What conditions support these inequalities?
  - Condition 1:

$$\omega^* > \underline{\omega}$$
$$2\omega_i - 2c - \overline{\omega} > \underline{\omega}$$
$$2\omega_i - \overline{\omega} - \underline{\omega} > 2c$$
$$\frac{2\omega_i - \overline{\omega} - \underline{\omega}}{2} > c$$
$$\omega_i - \frac{\overline{\omega} + \underline{\omega}}{2} > c$$

  - Condition 2:

$$\omega^* < \overline{\omega}$$
$$2\omega_i - 2c - \overline{\omega} < \overline{\omega}$$
$$2\omega_i - 2\overline{\omega} < 2c$$

14

$$\omega_i - \overline{\omega} < c$$

- Condition 3:

$$e^* > 0$$

$$b(2\omega_i - 2c - \overline{\omega}) > 0$$

$$2\omega_i - 2c - \overline{\omega} > 0$$

$$2\omega_i - \overline{\omega} > 2c$$

$$\frac{2\omega_i - \overline{\omega}}{2} > c$$

$$\omega_i - \frac{\overline{\omega}}{2} > c$$

- The only binding constraint is:

$$c < \omega_i - \frac{\overline{\omega} + \underline{\omega}}{2}$$

  - Intuition: If the cost of giving an unfavorable recommendation is too large, the merit commission will never be willing to give an unfavorable recommendation, even to low-competence judges.

- To confirm this solution, we can check for profitable deviations by the judge. A type $\omega < \omega^*$ that plays $\underline{e}^* = 0$ in equilibrium will never deviate to any $0 < e' < \overline{e}^*$ because their utility would be strictly decreasing. The cost of effort would be increasing and they would still not receive b. A type $\omega \geq \omega^*$ that plays $\overline{e}^*$ would never deviate to any $e' > \overline{e}^*$, because their utility would be strictly decreasing. They already receive b and the cost of effort would be increasing. If a type $\omega \geq \omega^*$ did deviate from $\overline{e}^*$, it would be to $e' = 0$, the minimum value of e, as their utility is strictly decreasing in e.

- Thus, the only deviations that we have to prevent are a type $\omega < \omega^*$ deviating from $\underline{e}^* = 0$ to $\overline{e}^*$ and a type $\omega \geq \omega^*$ deviating from $\overline{e}^*$ to $\underline{e}^* = 0$. The equilibrium solutions for the cut-points prevent these deviations.

- First, we can check whether a type $\omega^* + \epsilon > \omega^*$, who plays $\overline{e}^*$ in equilibrium, has an incentive to deviate to $\underline{e}^*$:

$$u_{\omega^*+\epsilon}(\overline{e}^*) \geq u_{\omega^*+\epsilon}(\underline{e}^*)$$

$$b - \frac{\overline{e}^*}{\omega^* + \epsilon} \geq 0$$

$$b \geq \frac{\overline{e}^*}{\omega^* + \epsilon}$$

$$b(\omega^* + \epsilon) \geq \overline{e}^*$$

$$b(2\omega_i - 2c - \overline{\omega} + \epsilon) \geq b(2\omega_i - 2c - \overline{\omega})$$

$$b\epsilon \geq 0$$

- This inequality is true by assumption, so there is no profitable deviation.
- Second, we can check whether a type $\omega^* - \epsilon < \omega^*$, who plays $\underline{e}^*$ in equilibrium, has an incentive to deviate to $\overline{e}^*$:

$$u_{\omega^* - \epsilon}(\underline{e}^*) \geq u_{\omega^* - \epsilon}(\overline{e}^*)$$

$$0 \geq b - \frac{\overline{e}^*}{\omega^* - \epsilon}$$

$$\frac{\overline{e}^*}{\omega^* - \epsilon} \geq b$$

$$\overline{e}^* \geq b(\omega^* - \epsilon)$$

$$b(2\omega_i - 2c - \overline{\omega}) \geq b(2\omega_i - 2c - \overline{\omega} - \epsilon)$$

$$0 \geq -b\epsilon$$

$$b\epsilon \geq 0$$

- This inequality is true by assumption, so there is no profitable deviation.
- Step 4: Are the conditions that support the sequential rationality of the on-path actions of the merit commission compatible with the equilibrium cut-points on the type of the judge, ω^*, and on the level of effort chosen by the judge, e^*?
  - Condition 1:

$$c \geq \frac{2\omega_i - \omega^* - \overline{\omega}}{2}$$

16

$$2c \geq 2\omega_i - \omega^* - \overline{\omega}$$

$$2c \geq 2\omega_i - (2\omega_i - 2c - \overline{\omega}) - \overline{\omega}$$

$$2c \geq 2\omega_i - 2\omega_i + 2c + \overline{\omega} - \overline{\omega}$$

$$0 \geq 0$$

- Condition 2:

$$c \leq \frac{2\omega_i - \omega^* - \underline{\omega}}{2}$$

$$2c \leq 2\omega_i - \omega^* - \underline{\omega}$$

$$2c \leq 2\omega_i - (2\omega_i - 2c - \overline{\omega}) - \underline{\omega}$$

$$2c \leq 2\omega_i - 2\omega_i + 2c + \overline{\omega} - \underline{\omega}$$

$$0 \leq \overline{\omega} - \underline{\omega}$$

$$\underline{\omega} \leq \overline{\omega}$$

- The equilibrium solutions for ω^* and e^* are always compatible with the conditions that support the sequential rationality of the on-path actions for the merit commission by assumption.

- **Proposition 1:** There exists a unique semi-separating perfect Bayesian equilibrium (PBE) with two on-path signals of the form $\{\sigma_{J|\omega}(\sigma_C), \sigma_C(\sigma_{J|\omega}), \mu_C(\omega \mid e)\}$, where $\sigma_{J|\omega}(\sigma_C)$ is the strategy profile of the judge, $\sigma_C(\sigma_{J|\omega})$ is the strategy profile of the merit commission, and $\mu_C(\omega \mid x)$ are the beliefs of the merit commission. The strategies and beliefs of the players in equilibrium are as follows:

$$\sigma_{J|\omega}(\sigma_C) = \begin{cases} e = \overline{e}^* & \text{if } \omega \geq \omega^* \\ e = \underline{e}^* & \text{if } \omega < \omega^* \end{cases}$$

$$\sigma_C(\sigma_{J|\omega}) = \begin{cases} r = F & \text{if } e \geq \overline{e}^* \\ r = U & \text{if } e < \underline{e}^* \end{cases}$$

$$\mu_C(\omega \mid e) = \begin{cases} \omega \sim \mathcal{U}(\omega^*, \overline{\omega}) & \text{if } e = \overline{e}^* \\ \omega \sim \mathcal{U}(\underline{\omega}, \omega^*) & \text{if } e = \underline{e}^*, \\ \omega \sim \mathcal{U}(\underline{\omega}', \overline{\omega}') & \text{if } \underline{e}^* < e' < \overline{e}^*, \\ \omega \sim \mathcal{U}(\underline{\omega}'', \overline{\omega}'') & \text{if } e'' > \overline{e}^*, \end{cases}$$

where

$$\underline{e}^* = 0$$

$$\overline{e}^* = b(2\omega_i - 2c - \overline{\omega})$$

$$\omega^* = \frac{\overline{e}^*}{b} = 2\omega_i - 2c - \overline{\omega}.$$

This equilibrium exists if and only if:

$$c < \omega_i - \frac{\overline{\omega} + \underline{\omega}}{2}$$

$$\underline{\omega}'' + \overline{\omega}'' > \underline{\omega}' + \overline{\omega}'.$$

Comparative statics

- **Result 1:** In equilibrium, the effort expended by the judge to satisfy the criteria of the merit commission, $\overline{e}^*$, is increasing in the benefit to the judge of being reappointed. As b increases, $\overline{e}^*$ increases.

  - Proof:

  $$\frac{\partial \overline{e}^*}{\partial b} = \frac{\partial}{\partial b} b(2\omega_i - 2c - \overline{\omega})$$

  $$\frac{\partial \overline{e}^*}{\partial b} = 2\omega_i - 2c - \overline{\omega}$$

  $$\frac{\partial \overline{e}^*}{\partial b} > 0$$

  - Note: We know that $2\omega_i - 2c - \overline{\omega} > 0$ because this is the value of $\omega^* > 0$.

- **Result 2:** In equilibrium, the effort expended by the judge to satisfy the criteria of the merit commission, $\overline{e}^*$, is increasing in the minimum standard of competence demanded by the merit commission, ω_i. As ω_i increases, $\overline{e}^*$ increases.

18

- Proof:

$$\frac{\partial \overline{e}^*}{\partial \omega_i} = \frac{\partial}{\partial \omega_i} b(2\omega_i - 2c - \overline{\omega})$$

$$\frac{\partial \overline{e}^*}{\partial \omega_i} = 2b$$

$$\frac{\partial \overline{e}^*}{\partial \omega_i} > 0$$

- **Result 3:** In equilibrium, the effort expended by the judge to satisfy the criteria of the merit commission, $\overline{e}^*$, is decreasing in the cost of giving an unfavorable recommendation, c. As c increases, $\overline{e}^*$ decreases.

  - Proof:

$$\frac{\partial \overline{e}^*}{\partial c} = \frac{\partial}{\partial c} b(2\omega_i - 2c - \overline{\omega})$$

$$\frac{\partial \overline{e}^*}{\partial c} = -2b$$

$$\frac{\partial \overline{e}^*}{\partial c} < 0$$

- **Result 4:** In equilibrium, the effort expended by the judge to satisfy the criteria of the merit commission, $\overline{e}^*$, is decreasing in the upper bound of the distribution of the judge's type, $\overline{\omega}$. Increasing $\overline{\omega}$ increases the expected competence of the judge. As $\overline{\omega}$ increases, $\overline{e}^*$ decreases.

  - Proof:

$$\frac{\partial \overline{e}^*}{\partial \overline{\omega}} = \frac{\partial}{\partial \omega_i} b(2\omega_i - 2c - \overline{\omega})$$

$$\frac{\partial \overline{e}^*}{\partial \overline{\omega}} = -b$$

$$\frac{\partial \overline{e}^*}{\partial \overline{\omega}} < 0$$