AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Applied Psychology

ISSN: 0021-9010

https://doi.org/10.1037/apl0001064

Motivational Drivers for Serial Position Effects: Evidence From High-Stakes Legal Decisions

Ori Plonsky[1], Daniel L. Chen[2], Liat Netzer[3], Talya Steiner[4], and Yuval Feldman[5, 6]

[1] Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology

[2] Toulouse School of Economics

[3] aChord Center, The Hebrew University of Jerusalem

[4] Faculty of Law, The Hebrew University of Jerusalem

[5] Faculty of Law, Bar Ilan University

[6] Israel Democracy Institute, Jerusalem, Israel

Experts and employees in many domains make multiple similar but independent decisions in sequence. Often, the serial position of the case in the sequence influences the decision. Explanations for these serial position effects focus on the role of decision-makers' fatigue, but these effects emerge also when fatigue is unlikely. Here, we suggest that serial position effects can emerge due to decision-makers' motivation to be or appear consistent. For example, to avoid having inconsistencies revealed, decisions may become more favorable toward the side that is more likely to put a decision under scrutiny. As a context, we focus on the legal domain in which many high-stakes decisions are made in sequence and in which there are clear institutional processes of decision scrutiny. We analyze two field data sets: 386,109 U.S. immigration judges' decisions on asylum requests and 20,796 jury decisions in 18th century London criminal court. We distinguish between five mechanisms that can drive serial position effects and examine their predictions in these settings. We find that consistent with motivation-based explanations of serial position effects, but inconsistent with fatigue-based explanations, decisions become more lenient as a function of serial position, and the effect persists over breaks. We further find, as is predicted by motivational accounts, that the leniency effect is stronger among more experienced decision-makers. By elucidating the different drivers of serial position effects, our investigation clarifies why they are common, when they are expected, and how to reduce them.

Keywords: sequential decision-making, judicial decision-making, legal decisions, order effects, consistency

Supplemental materials: https://doi.org/10.1037/apl0001064.supp

In many contexts, professionals are required to make decisions concerning similar but completely independent cases in sequence. Parole judges grant or deny parole, quality assurance specialists inspect products, and analysts produce forecasts. In these and in similar professional contexts, cases are presented in sequence for mere practical or operational reasons, and the sequential structure should have no bearing on decisions. However, decisions made in sequence often differ from those made in isolation (e.g., Bhargava & Fisman, 2014; Chen et al., 2016; Cohen et al., 2020; Hartzmark & Shue, 2018; Leibovitch, 2016; Read & Loewenstein, 1995; Simonsohn & Gino, 2013; Simonson, 1990; Stewart, 2009; Sunstein et al., 2001). One factor that past studies have shown could potentially—and undesirably—impact decisions concerning completely independent cases in a sequence is the serial position of the case.

Why does the serial position of a case influence decisions? Past research documenting serial position effects mostly focuses on the role of decision-makers' fatigue in higher serial positions. For example, it has been argued that accumulating fatigue over a sequence of independent cases leads parole judges to be increasingly more harsh as they make more decisions concerning inmates (Danziger et al., 2011), food health inspectors to be increasingly

Ori Plonsky https://orcid.org/0000-0002-1619-4482

Yuval Feldman https://orcid.org/0000-0001-8416-865X

The authors thank Andreas Glöckner for valuable discussions on an earlier version of this article. This research was supported by the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013), ERC Grant 324182. Daniel L. Chen acknowledges Institute for Advanced Study in Toulouse funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, Grant ANR-17-EUR-0010, and financial support from the research foundation Toulouse School of Economics Partnership and Artificial and Natural Intelligence Toulouse Institute funding.

Earlier versions of this article were posted on the Social Science Research Network (https://ssrn.com/abstract=3414155). Parts of this article were presented in the following conferences: 2019 Public Policy Conference in the Israel Democracy Institute; Subjective Probability, Utility, and Decision Making 2019; The 5th Coller Conference on Behavioral Economics; and the 42nd Annual Meeting of the Society for Judgement and Decision Making (as the Einhorn Award Lecture).

Correspondence concerning this article should be addressed to Ori Plonsky, Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology, Haifa 3200003, Israel. Email: oplonsky@gmail.com

less stringent as they visit more food establishments (Ibanez & Toffel, 2020), and journal editors to be increasingly more positive as they decide concerning more initial submissions (Orazbayev, 2017).[1]

Although fatigue can drive serial position effects, we argue that they often emerge independently of fatigue. We suggest that the motivation of decision-makers to be consistent across and within sequences of decisions (Falk & Zimmermann, 2017; Haubensak, 1992) can be an important driver of many real-world serial position effects. Motivational accounts of serial position effects are rarely considered in organizational contexts, but they can have important practical consequences. For example, serial position effects that are mainly driven by fatigue can be mitigated by adding breaks between different cases in a sequence, but a similar intervention will not work if the effects are driven by decision-makers' motivations. Moreover, as we argue below, there are reasons to believe that experienced decision-makers would show stronger serial position effects when these are driven by motivation (but not when they are driven by fatigue).

To demonstrate the role of motivational accounts in driving serial position effects, this article analyzes sequential decisions in U.S. immigration courts and in 18th century "Old Bailey," London's main criminal court. In both these contexts, fatigue and motivational mechanisms are expected to trigger serial position effects in opposite directions. This makes the empirical investigation of these high-stakes contexts suitable for testing whether motivational accounts contribute to the emergence of such effects. To foreshadow our results, we find that in both settings, serial position effects emerge in the direction predicted by motivational, but not fatigue-based, accounts. Before presenting our empirical setting and analysis, we provide a theoretical overview of the different channels that can lead to serial position effects in sequential decisions.

Potential Mechanisms of Serial Position Effects

Fatigue-Based Mechanisms

Fatigue is known to influence cognitive functions in many ways (Hockey, 2013). It increases distractibility, changes information processing, lowers motivation and engagement, alters metacognitive processes, increases likelihood of errors, and affects mood (e.g., Bartlett, 1943; Boksem et al., 2005; Deng et al., 2018; Lorist et al., 2005; Shockley et al., 2021; Sonnentag & Zijlstra, 2006; van der Linden et al., 2003). In turn, these effects can lead to systematic differences in decision-making.

Most past research on serial position effects in sequences of completely independent cases assumes they are driven by fatigue. In particular, two main fatigue-based mechanisms have been discussed (although not clearly distinguished, see below): increased choice of the status quo/default option as a result of mental depletion and increased choice of the option that elicits less time and effort. We begin our theoretical analysis by considering these two mechanisms.

Status Quo/Default

It has been suggested that the more decisions one has to make in a session, the more likely he or she is to make an easier choice by accepting the status quo or the default option. This theory relies on the premises that decision-making depletes mental resources and that overruling a default and/or changing the status quo depletes more mental resources than preserving them (Levav et al., 2010; Polman & Vohs, 2016; Vohs et al., 2008). Increase in choice of the status quo, default option was proposed as the underlying mechanism in one of the most high-profile studies documenting serial position effects in the field. Studying Israeli parole judges, Danziger et al. (2011) found that the chances of a judge to grant parole in a hearing dramatically decreased as a function of the serial position of the hearing in a session. In support of the role of fatigue in driving this effect, the chances of parole were restored to their initial levels at the beginning of the next session, even though sessions were divided only by short food breaks. The authors then suggested that tired and hungry judges prefer to accept the default, status quo option, which they assumed was to deny a prisoner's request for parole. Later works, however, criticized this interpretation of the data, claiming that unobservable features unrelated to fatigue may (at least partially) drive the observed results (Glöckner, 2016; Weinshall-Margel & Shapard, 2011) and that the proposed process of mental depletion should have manifested differently (Daljord et al., 2019).

Differences in Effort

Fatigue can also drive serial position effects via a more parsimonious mechanism that does not assume anything about decision-makers' mental states. It is commonly assumed that as people get more tired, they tend to select choice options that elicit less effort (Hull, 1943; Solomon, 1948). Many serial position effects may be the result of people getting increasingly likely to avoid decisions that elicit more time and effort. Hospital caregivers wash their hands less over the course of a shift (Dai et al., 2015); voters abstain more in contests appearing further down the ballot (Augenblick & Nicholson, 2016); inspectors of food establishments cite fewer health and safety violations later in their workday (Ibanez & Toffel, 2020); financial analysts update their forecasts less and issue less unique forecasts after making more forecasts in a day (Hirshleifer et al., 2019); and clinicians offer to patients less procedures later in the day (Hsiang et al., 2019; R. H. Kim, Day, et al., 2018; Persson et al., 2019). In support of the role of fatigue in driving some of the aforementioned patterns, breaks were found to moderate or even eliminate some of these serial position effects. Moreover, in some contexts, when choices required additional effort (e.g., when a longer inspection in a food establishment meant a prolonged inspector shift; Ibanez & Toffel, 2020), the effects were magnified.

Relationship Between the Two Mechanisms

Often, preserving the status quo (or accepting a default) requires less time and effort than other decision alternatives. This potentially confounds the two fatigue-related mechanisms as drivers of serial position effects. For example, in Israeli parole hearings, the decision

[1] A related, but distinct, line of research studies serial position effects in sequences of cases that are then explicitly compared and ranked, for example, to select a winner, as in contests (Antipov & Pokryshevskaya, 2017; Bian et al., 2022; Bruine de Bruin, 2005, 2006; Glejser & Heyndels, 2001; Haan et al., 2005; B. Kim et al., 2021; Li & Epley, 2009; Mantonakis et al., 2009; Rotthoff, 2015; Scheer, 1973; Wilson, 1977). In such contexts, evaluations are not likely to be truly independent of one another (even when they are formally supposed to be). We return to this line of research when discussing the calibration mechanism below, but the focus in this article is on contexts in which cases in the sequence are never explicitly ranked, and decisions are made in sequence for convenience only.

to grant parole, which presumably overrules the status quo, entails an additional discussion of the conditions set for the paroled inmate (N. Dagan, personal communication, June 2020). That is, denial of parole is both the status quo option and the option eliciting less effort, and thus the increase in denials of parole is consistent with either fatigue mechanism. Similar arguments can be made for overlooking safety violations in health establishments, for abstaining in voting contests, and for updating financial forecasts. To the best of our knowledge, there is no clear evidence of serial position effects that are plausibly driven by one of the two fatigue mechanisms, but not by the other. Yet, in principle, the two mechanisms provide different predictions in some contexts. For example, it could be that while one decision is more effortful than another, neither is a default and both actively change the status quo.

Motivational Accounts

Some serial position effects cannot be explained by fatigue. For example, studies documented serial position effects that persist over long overnight breaks (e.g., Colton & Peterson, 1967; Glejser & Heyndels, 2001) and that emerge in simulated settings over short time spans in which fatigue is unlikely to have been accumulated (e.g., Unkelbach et al., 2012; Unkelbach & Memmert, 2008). To explain serial position effects when fatigue is a less plausible driver, several studies (see review in Unkelbach & Memmert, 2014) have proposed that they emerge because decision-makers strive to produce internally consistent sequences of decisions by strategically avoiding extreme judgments early in the sequence. This mechanism, called calibration, was mainly proposed in contexts that are quite different from those we focus on in this article; nevertheless, the premise that decision-makers are motivated to make consistent decisions can help explain many real-world serial position effects.

Motivation for consistency may stem from a direct preference for consistency, for example, so as to avoid cognitive dissonance (Yariv, 2002), or from the fact that consistency violations are commonly perceived as signals of low skill and reliability (Falk & Zimmermann, 2017). We propose that the motivation to be or appear consistent can manifest in several ways, leading to serial position effects based on three distinct mechanisms: calibration (mentioned above) that leads to increases in extreme decisions, quotas that lead to increases in decisions restricted under quota, and strategic uncertainty that leads to increases in decisions that are likely to be scrutinized. We elaborate on each of them next.

Calibration

Decision-makers who make sequences of categorical decisions need to use a mapping from item attributes to the categorical scale (e.g., lenient vs. harsh decisions). It has long been argued that this mapping should be calibrated according to the expected range of input levels (Parducci, 1965), such that the most extreme inputs will be mapped to the endpoints of the scale. Past research (Fasold et al., 2012; Unkelbach et al., 2012; Unkelbach & Memmert, 2014) suggested that because decision-makers strive to maintain internal consistency in their sequential evaluations (Haubensak, 1992), and because they may initially have some uncertainty concerning the expected range of input levels, they will likely avoid extreme judgments early in the sequence (in the "calibration" stage) to preserve the freedom to make more extreme judgments later on. Avoiding extreme judgments early in the sequence, but not later, then leads to an increase in extreme decisions with serial position.

The process of calibration is particularly necessary when evaluations of items in the sequence are later used to create a ranking of cases. For example, in sports contests, judges that score each performance on a scale of up to 10 may be initially unsure what level of performance deserves a 10 and thus avoid assigning this extreme rating so that it will be possible to later give higher scores to better performances and preserve internal consistency. Indeed, in real-world contexts, the calibration explanation for serial position effects was primarily suggested in contest settings (Antipov & Pokryshevskaya, 2017; Bian et al., 2022; Bruine de Bruin, 2005; Rotthoff, 2015; but see Memmert et al., 2008). In such settings, unlike those that we consider in this article, an implicit comparison among items in the sequence is almost inevitable, and evaluations are likely not (and arguably need not be) truly independent. However, decision-makers may strive to produce internally consistent sequences of decisions even when the items in the sequence are not later compared and ranked (Falk & Zimmermann, 2017; Haubensak, 1992). Thus, as long as decision-makers need to calibrate an evaluation scale (i.e., they have some initial uncertainty concerning the range of quality of items in the sequence), they may still strategically avoid early extreme judgments, making extreme decisions more likely in later serial positions.

Two points are notable. First, this mechanism relies on early calibration of the input–output function. Experienced or expert evaluators should presumably start the sequence already well-calibrated and thus should be less affected by the need for calibration (Bian et al., 2022). Yet, experts are also more likely to be concerned with consistency violations than novices and more experienced professionals may have learned the importance of maintaining degrees of freedom for later evaluations (Unkelbach et al., 2012; Unkelbach & Memmert, 2014). Indeed, in one experimental demonstration of calibration effects, both novices and experts exhibited similarly strong serial position effects, and the effect was even directionally larger with experts (Unkelbach et al., 2012). Second, calibration effects should emerge only if the decision-maker anticipates that more than one evaluation will be made. Otherwise, it is pointless to strategically avoid an extreme judgment in the first (and only) evaluation. Hence, on average, judgments on cases that are not followed by other cases are expected to be more extreme than if the same case would have been presented as the first case in a sequence with multiple cases (Fasold et al., 2012). We return to this prediction when testing the mechanism in one of our empirical contexts.

Quotas

Previous research suggests that decision-makers may have implicit quotas (either external or self-imposed) on the number or proportion of certain types of decisions they make in a sequence (Bhargava & Fisman, 2014; Chen et al., 2016). Self-imposed quotas may form because people strive to produce sequences of decisions that are consistent with what they, or others, think is reasonable or expected (Simonsohn & Gino, 2013). For example, if in a certain context historically about 25% of decisions in each sequence were lenient (and the others harsh), a decision-maker may be reluctant to deviate much from this 25% mark and thus impose a quota for lenient decisions.

How might having a quota imply serial position effects? We suggest that to increase the chances that an ongoing sequence will eventually adhere to an imposed quota, a decision-maker continuously monitors how much of the quota has been "used" so far. The more of it has already been used, the more hesitant the decision-maker would be to use more of it. For example, under a quota for lenient decisions, the likelihood of a lenient decision in a case should be smaller if the case follows several lenient decisions (i.e., much of the quota has been used) than if it follows several harsh decisions. Put differently, a decision-maker who is restricted under quota should dynamically adjust a "decision criterion" such that the more decisions restricted under quota have been made so far, the stricter the criterion (and vice versa). The restricted decision is then given only to cases that achieve the updated criterion.

The dynamics of changing a "decision criterion" according to the history of decisions in the sequence only implies serial position effects with the additional (untested) assumption that *initially* the criterion is set relatively strictly (a restricted decision is not given easily). This assumption implies that early items are less likely to achieve the criterion (than had they not been subject to quotas), and all else equal, the criterion is more likely to be relaxed for ensuing cases. Hence, with this assumption, the quotas mechanism predicts that the decision restricted under quota will become more likely as a function of serial position. A strict initial criterion may be reasonable because the decision-maker would be wise to hedge against the possibility that many cases will require using the quota even if the criterion is raised.

This dynamic process (even without the "initial strictness" assumption) has two main testable implications. First, it implies negative autocorrelation between consecutive decisions: A lenient decision in one case makes the decision criterion stricter and thus increases the likelihood that the next decision will be harsh (and vice versa). Second, because the decision criterion depends on the proportion of the decisions restricted under quota that were made in the sequence so far, holding this proportion fixed implies conditional independence of decisions (and elimination of serial position effects that are driven by this mechanism). We will use these properties when testing the mechanism in our empirical contexts.

Strategic Consistency

Decision-makers strive to make decisions that are both accurate and consistent over time, but accuracy and consistency may sometimes conflict (an accurate decision may be inconsistent with a past decision in a similar case). We posit that when such conflicts emerge, either decision could be perceived by others as a sign of bias or incompetence, and experienced decision-makers are aware of this problem. We suggest that to avoid having others reveal their inconsistency (or inaccuracy), decision-makers then tend to make the decision that is *less likely to be scrutinized* by others (if there is one). That is, because in the presence of a consistency–accuracy conflict, any decision can reflect poorly on decision-makers, they tend to choose in a way that increases the chances that it will not come under scrutiny at all. The more decisions one makes in a sequence, the more likely it is that two cases will be sufficiently similar so that consistency considerations will matter, so that decisions in later cases are more likely to be subject to consistency–accuracy conflicts. Thus, a tendency toward decisions that are less likely to be scrutinized becomes increasingly more likely in later cases of a sequence, a serial position effect.[2] A specific implication of this mechanism is that the serial position effect would depend on the decision-maker's expectation for the level of scrutiny that a particular decision will draw: It is likely to be stronger when decision scrutiny is more likely.

Relationship Among the Motivational Mechanisms

The three motivational mechanisms all share the basic assumption that serial position effects emerge due to decision-makers' aim to preserve consistency in sequential decisions: Calibration effects are driven by the desire to produce internally consistent sequences, quota effects are driven by the desire to produce sequences consistent with historical averages or an expected sequence, and strategic consistency effects are driven by a desire not to have consistency (or accuracy) violations revealed by others. Furthermore, the implied predictions of the three mechanisms regarding the direction of serial position effects may also often converge. In many settings, decisions that can be considered more extreme or atypical (and thus predicted by the calibration account to get increasingly more likely) will also be those more likely to be restricted under quota (and thus predicted by the quotas mechanism to get increasingly more likely), and possibly also those more likely to later come under scrutiny (and thus predicted by the strategic consistency account to get increasingly more likely).

Nevertheless, the three motivational mechanisms differ in their underlying assumptions, and it should thus be possible to distinguish among them. For example, the quotas mechanism predicts negative autocorrelation between consecutive decisions that the other two mechanisms do not, whereas the strategic consistency mechanism predicts that a serial position effect would emerge only when one type of decision is much less likely to be scrutinized. Hence, both statistical analysis and analysis of the institutional details of a particular decision context can help disentangle the different mechanisms underlying a particular serial position effect.

Summary of Mechanism Predictions and Predicted Moderators

To round off our theoretical section, we present the main generic predictions of the five mechanisms we identified as drivers of serial position effects and develop two additional predictions concerning potential moderators of these effects. As summarized in Table 1, two of the five potential mechanisms involve the effects of fatigue: mental depletion that leads to increase in propensity to choose the status quo/default option and fatigue that leads to increase in propensity to choose the less effortful option. Three other mechanisms involve decision-makers' motivations to be consistent: calibration that predicts an increase in the propensity to choose an extreme option, quotas that predicts an increase in the propensity to choose the option that is limited under quota, and strategic consistency that predicts an increase in the propensity to choose the option more likely to be scrutinized.

[2] Avoiding decisions that come under scrutiny may also influence all decisions regardless of serial position and regardless of consistency considerations. Yet, to explain why avoidance of scrutiny increases with serial position, the additional assumption of consistency is vital.

Table 1
Summary of Potential Mechanisms Driving Serial Position Effects

	Predicted association with serial position		
		Implied prediction	
Mechanism	Generic prediction: Increased choice of …	In Study 1	In Study 2
Fatigue-based			
Status quo/default	The status quo/default option	Harshness	Harshness
Differences in effort	The less effortful option	No effect	No effect
Motivation-based			
Quotas	The option limited by a quota	Leniency	Leniency
Calibration	The more extreme option(s)	Leniency	Leniency
Strategic consistency	The option that is scrutinized less	Leniency	Leniency

Note. Derivations of the directional predictions for Study 1 and Study 2 are made within those studies below. The predictions are presented here for expositional purposes.

In addition to the specific generic predictions of each mechanism, the fundamental differences between fatigue and motivation for consistency as classes of explanations for serial position effects give rise to predictions concerning two potential moderators of these effects: the possibility to take a break between decisions and the decision-makers' experience. Breaks and diminished load allow people to recover from fatigue (Hockey & Earle, 2006; Hunter & Wu, 2016; S. Kim, Park, et al., 2018; Sievertsen et al., 2016; Trougakos & Hideg, 2009). Hence, when fatigue underlies the serial position effects, reducing decision load and increasing break times between consecutive decisions should reduce or even eliminate these effects (with longer breaks more likely to eliminate them). However, when serial position effects are triggered by decision-makers' motivations, adding breaks is unlikely to have much effect. Thus, breaks between sequential decisions should moderate the effects of serial position triggered by fatigue, but not those triggered by motivation for consistency.

The moderating role of experience and expertise is different. Motivation to produce consistent sets of decisions is probably more prevalent among expert and experienced decision-makers who should be more aware of the importance of producing consistent decisions and more affected by the consequences when they do not preserve consistency. Indeed, research suggests that more experienced managers tend to be more consistent with previous actions and particularly when actions are made explicitly and publicly (Finkelstein & Hambrick, 1990). Moreover, research also suggests that the higher the degree of expertise, the lower is the perceived task variability (Haerem & Rau, 2007). Thus, more experienced decision-makers may be more likely to perceive two items in a sequence as sufficiently similar so that they would merit consistent decisions. Finally, consistency in treatment of similar stimuli is considered a necessary condition for expertise (Einhorn, 1974; Weiss & Shanteau, 2003). Serial position effects triggered by motivational accounts should thus be magnified (or more likely to exist) among expert and more experienced decision-makers.

In contrast, serial position effects driven by fatigue should, if anything, be *weaker* among experienced decision-makers. Research suggests that experienced decision-makers use less cognitive resources to make quicker and more intuitive decisions (Salas et al., 2010) and are better at avoiding problems associated with information overload (Shanteau, 1995), implying they would accumulate less fatigue with each additional decision. Moreover, experienced decision-makers may be more able to adjust for the impacts of fatigue on their decisions (Hirshleifer et al., 2019). Indeed, research suggests that induced fatigue impacts the performance of professional athletes less than the performance of novices (Lyons et al., 2006). Hence, whereas the motivation-based mechanisms predict that more experienced decision-makers would exhibit a stronger serial position effect, the fatigue-based mechanisms do not.

The Present Research

The primary goal of the present study is to highlight the role of motivational accounts in driving serial position effects in sequential decision-making. To that end, we investigate important real-world contexts in which all three motivation-based mechanisms predict serial position effects that cannot be predicted by the two fatigue-based mechanisms. We aim to demonstrate that, consistent with motivational accounts, these serial position effects emerge, are not eliminated by breaks, and are moderated by decision-makers' experience.

We empirically test for serial position effects in two very high-stakes sequential decisions settings from the legal domain: decisions concerning asylum applications in U.S. immigration courts (Study 1) and criminal case verdicts of juries in 18th century London criminal court (Study 2). The focus on the legal domain has two main advantages. First, in many legal contexts, it can be straightforward to derive the directional predictions of the potential drivers of serial position effects and thus to empirically evaluate contrasting predictions. Specifically, many legal decisions concern the question of whether to preserve or change the status quo (e.g., keeping the prisoner in jail or upholding the decision of another court), and in many legal contexts, there often exists a formal system of decision scrutiny (e.g., appeals). Thus, the predictions of the default/status quo and of the strategic consistency mechanisms should be evident.

Second, a famous study of serial position effects (Danziger et al., 2011), often cited in the literature as a prime example of the impact of fatigue on high-stakes sequential decision-making, involved legal decisions. Interestingly, the context of that study, parole decisions in Israel, includes a relatively unique feature in legal settings: decisions are unlikely to be scrutinized as an appeal mechanism is virtually nonexistent (Assy & Menashe, 2014). Hence, the strategic consistency motivational mechanism, which predicts an increase in decisions that come under more scrutiny, has no directional predictions

in that context. Examining the relative importance of fatigue and motivational accounts in driving serial position effects in more typical legal contexts is thus valuable. To increase the likelihood that our findings will generalize well within the legal domain, we investigate two very distinct contexts. Study 1 involves short sequences of few individually made decisions that follow long hearings concerning hapless persons seeking refuge, whereas Study 2 involves long sequences of many group decisions that follow short hearings concerning suspected criminals. Converging results in these distinct contexts should help diminish concerns that an atypical context rather than the fatigue versus motivation distinction drives our findings.

In what follows, we start the discussion on each study with a derivation of the directional hypotheses for each of the five mechanisms. We then test these hypotheses by checking for the existence and direction of a serial position effect and the moderating effects of breaks and of experience. Finally, because we find that the results of those tests are consistent with motivational accounts, we present exploratory analyses investigating which of the three motivational explanations is most likely.

Study 1: Asylum Decisions in Immigration Courts

In Study 1, we investigate the effect of serial position on sequential legal decisions in a context with very high stakes: applications for asylum in U.S. immigration courts. We refer to granted asylum applications as lenient decisions and to denials of applications as harsh. We use administrative data on U.S. refugee asylum cases adjudicated in immigration courts between 1980 and 2013. Our sample includes 386,109 cases that were adjudicated by 425 judges with an average of 8.54 years of experience since appointment (SD = 6.29) in 53 immigration courts nationwide. Each immigration court covers a geographic region, and within each court, cases are randomly assigned to judges using a computerized algorithm. Cases are scheduled on a judge's docket for the next available date and time, which are often several months in advance. Judges very rarely reshuffle the order of scheduled hearings. The number of asylum hearings within a judge's day varies between one and five (M = 1.51). Supplemental Tables S1–S3 provide additional summary statistics. We begin by examining the predictions of each of the five mechanisms presented above in this context.

Derivation of Mechanism Predictions

Status Quo/Default (Increased Choice of the Status Quo/Default)

The decisions we analyze concern cases in which the applicant asks to be granted a new status that he or she does not currently hold, much like a prisoner filing for parole. Hearings of asylum cases are adversarial. An attorney from the Department of Homeland Services (DHS) argues that asylum is unwarranted, and the asylum seeker must prove he or she has a well-founded fear of persecution if deported. The burden of proof therefore lies on the applicant. Cases are of two main types. Affirmative cases are of applications that were previously denied by a professional asylum officer who reviewed the case. In defensive cases, the asylum seeker was apprehended by the DHS and found to be illegally present in the United States before he or she filed the application for asylum. In these cases, the applicant is usually held in a detention facility at the time of the hearing and remains detained until the end of proceedings. For all of these reasons, we assume *denial* of the application is probably considered the status quo, default, easier decision. Therefore, if default decisions are more likely in cases with higher serial positions, we would expect more asylum decisions to be harsh (denials) as a function of serial position.

Differences in Effort (Increased Choice of the Less Effortful Decision)

There does not seem to be a meaningful difference in the mean adjudication time of lenient and harsh decisions in this context. Although our data do not include the length of adjudication, we approximate the length of each hearing (that is not the last in a judge's day) by the difference between the start time of that hearing and the start time of the hearing that follows it. The difference between the mean approximated adjudication time of cases that were denied asylum (M = 2.98 hr, SD = 1.64) and that of cases that were granted asylum (M = 2.95 hr, SD = 1.68) is just 1.7 min, less than 1% of the total average adjudication time. Although the (approximate) length of adjudication provides no evidence that either lenient or harsh decisions elicit more effort, it is nonetheless possible that these decisions differ in effort in ways not reflected by adjudication time. If they do, then we may expect that in later cases, immigration judges will behave similarly to judges under high caseload. Under high caseload (see review in Engel & Weinshall-Margel, 2020), judges often rule more harshly and use more heuristics, and a common heuristic in judicial decisions is "passing the buck" (Dhami, 2003), which in our setting implies harsh decisions. Therefore, if anything, the current mechanism would predict more asylum decisions to be harsh as a function of serial position. Yet, because the institutional details do not clearly suggest that this is the case, to be conservative, our analysis assumes that this mechanism predicts no serial position effects.

Quotas (Increased Choice of the Decision Restricted Under Quota)

Immigration judges have considerable discretion over their decisions, and congruently, there is a very large heterogeneity in grant rates across different immigration judges. For example, even holding the nationality of the applicant constant, the grant rates of different judges in the same court at the same time window can range between 5% for the harshest judge and 96% for the most lenient one (Ramji-Nogales et al., 2007). Such discretion implies no external quotas are enforced on judges. It is hard to know a priori whether judges self-impose quotas (in the Results section, we present statistical evidence against this). Yet, if they do, it is more reasonable to assume that the decision restricted under quota is the lenient one (grant asylum). As discussed above, the easier or default decision is likely a denial of the asylum application, and in our data, most applications (66%) are denied. Hence, the likely prediction of an implicit (self-imposed) quota mechanism is increased leniency with serial position.

Calibration (Increased Choice of the More Extreme Decision)

For similar reasons discussed when analyzing the likely status quo/default option and the likely decision that may be restricted by

quotas, it seems reasonable to assume that the more extreme decision is a lenient one. Hence, as with self-imposed implicit quotas, calibration implies increased prevalence of lenient decisions as a function of serial position.

Strategic Consistency (Increased Choice of the Decision Less Likely to Be Scrutinized)

When an immigration judge denies an asylum request (i.e., issues a harsh ruling), the asylum seeker may appeal the decision to the Board of Immigration Appeals (BIA). In a recent study, Ash et al. (2020) found that 67% of denials are appealed. Moreover, using a machine learning approach, they estimated that denials that were not appealed in practice were five times less likely to have a successful appeal had they been appealed. This suggests that denials that have a chance to be overturned are regularly appealed to the BIA. In contrast, while the government may also appeal an approved asylum request, only 3.2% of the appeals to BIA were by the government. Together, these findings imply that harsh decisions are far more likely to be scrutinized than lenient decisions, and experienced judges are likely aware of this fact. Thus, if strategic consistency is at play, we would expect more lenient decisions as a function of serial position.

Analysis Strategy

As summarized in Table 1, all three motivational mechanisms of serial position effects predict increasing leniency (grant asylum) as a function of serial position. In contrast, neither fatigue-based mechanism predicts increasing leniency. Our analysis starts with a confirmation of the predictions made by the motivational mechanisms. To further establish a rejection of the fatigue-based mechanisms, we then show that increasing leniency with serial position persists even over (long) breaks and that this effect is stronger for more experienced judges, providing further support to the claim that the effect is driven by judges' motivations rather than fatigue. Finally, we try to identify which of the three motivational mechanisms is most likely to drive the effect in this case and find the data mostly support the strategic consistency account.

Our primary outcome variable is a binary indicator recording whether a lenient decision, asylum grant, was issued. For our main statistical analyses, we implement (using command *melogit* in Stata 15.1) a series of mixed-effects logistic regressions with crossed random intercepts for the judge and the court and with serial position of a case in a judge's day (or week) as the main independent variable. Parameter estimates are given using the Laplace approximation.

Transparency and Openness

The data analyzed in this study are not publicly available as it comes from the Transactional Records Access Clearinghouse (TRAC; https://trac.syr.edu/). Access to this data is restricted under the TRAC Fellows program and license agreement governing data use and is conditional on obtaining an appointment with TRAC. Researchers seeking to access the data may do so by contacting TRAC directly and undergo a standardized procedure. The analysis plan was not preregistered. We adhered to the *Journal of Applied Psychology* methodological checklist.

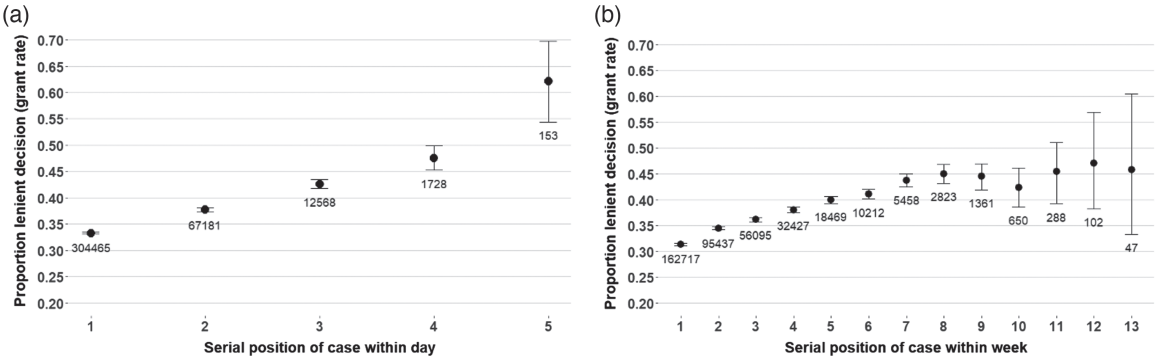
Results

We first investigate how the rate of lenient decisions—the proportion of approved asylum applications—changes as a function of the serial position of a case within a sequence of daily cases. Figure 1a shows a clear monotonic increasing trend. As most days include only one case, we repeat the analysis excluding days with only one case and find a similar trend (Supplemental Figure S1).

Table 2 presents the results of the mixed-effects regression models. Model 1 includes serial position as the only fixed effect. The coefficient for serial position is positive and significant ($\beta = 0.0523$, 95% CI [0.0381, 0.0665], $p < .001$, $OR = 1.054$), suggesting rulings in later serial positions tend to be more lenient. Model 2 adds controls for (a) the judge's prior experience and demographics, which have a large effect on the decisions on their

Figure 1

Proportion of Lenient Rulings (in Favor of an Asylum Applicant) by the Serial Position of a Case in a Sequence of Cases



Note. Unadjusted descriptive statistics. Numbers below the points indicate the number of cases that each point represents. Error bars indicate bootstrapped 95% confidence interval for the rates with 1,000 replicates. (a) Serial position within a day. Fourteen cases adjudicated after the fifth case are omitted. (b) Serial position within a week. Twenty-three cases adjudicated after the 13th case are omitted.

Table 2
Study 1 (Immigration Judges): Baseline Results, Daily Sequences

	Lenient decision dummy			
	Model 1	Model 2	Model 3	Model 4
Variable	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)
Serial position in the day	0.052*** (0.007)	0.101*** (0.011)	0.054*** (0.007)	0.101*** (0.011)
Day of the week			0.031*** (0.003)	0.042*** (0.003)
Republican judge		−0.090 (0.074)		−0.091 (0.074)
Male judge		−0.417*** (0.078)		−0.417*** (0.078)
Age of judge		0.035*** (0.001)		0.035*** (0.001)
Judge experience				
Over 8 years		−0.050*** (0.015)		−0.050*** (0.015)
Government		−0.101 (0.074)		−0.102 (0.074)
INS		−0.271*** (0.076)		−0.274*** (0.076)
Military		−0.379*** (0.111)		−0.384*** (0.112)
Private		−0.109 (0.076)		−0.113 (0.076)
Academia		0.068 (0.148)		0.069 (0.148)
Lawyer present		1.160*** (0.022)		1.163*** (0.022)
Defensive case		−0.501*** (0.009)		−0.502*** (0.009)
Applicant family size[a]				
2		0.254*** (0.015)		0.254*** (0.015)
3		0.353*** (0.025)		0.354*** (0.025)
4		0.504*** (0.032)		0.504*** (0.032)
5		0.505*** (0.056)		0.507*** (0.056)
6+		0.502*** (0.089)		0.502*** (0.089)
Applicant nationality[b]				
China		0.090*** (0.012)		0.090*** (0.012)
Haiti		−1.675*** (0.024)		−1.675*** (0.024)
El Salvador		−1.502*** (0.025)		−1.502*** (0.025)
Guatemala		−0.844*** (0.022)		−0.845*** (0.022)
Colombia		−0.482*** (0.022)		−0.482*** (0.022)
Number of cases in a day[a]				
2		−0.044*** (0.011)		−0.041*** (0.011)
3		−0.091*** (0.02)		−0.088*** (0.02)
4		−0.063 (0.038)		−0.058 (0.038)
5		0.232* (0.109)		0.242* (0.109)
6		1.183* (0.527)		1.214* (0.527)
Number of grant decisions of last five in court[a]				
1		0.144*** (0.014)		0.144*** (0.014)
2		0.219*** (0.014)		0.218*** (0.014)
3		0.317*** (0.015)		0.315*** (0.015)
4		0.470*** (0.018)		0.468*** (0.018)
5		0.636*** (0.025)		0.632*** (0.025)
Number of grant decisions of last five of judge[a]				
1		0.251*** (0.014)		0.252*** (0.014)
2		0.459*** (0.014)		0.46*** (0.014)
3		0.66*** (0.016)		0.661*** (0.016)
4		0.908*** (0.018)		0.909*** (0.018)
5		1.242*** (0.023)		1.245*** (0.023)
Constant	−1.17*** (0.116)	−3.280*** (0.141)	−1.26*** (0.116)	−3.421*** (0.142)
Random parts (*SD*s)				
Judge	0.870	0.638	0.871	0.640
Court	0.766	0.462	0.766	0.464
N	386,109	342,803	386,109	342,699

Note. Results of a mixed-effects logistic regression with random intercepts for judge and court. INS = Immigration and Naturalization Services; *SE* = standard error. Number of judges = 425 (Models 1, 3) or 358 (Models 2, 4). Number of courts = 53 (Models 1, 3) or 52 (Models 2, 4).
[a] Reference level for each of the ordinal variables is the minimal possible value (0 or 1). [b] Only the five most common nationalities were used, and the rest were classified as "other" (reference level).
* $p < .05$. *** $p < .001$.

own (cf. Ramji-Nogales et al., 2007); (b) observable case information, namely the presence of legal representation, the type of case (defensive or affirmative), the number of family members for which the application is made, and the nationality of the applicant; (c) the total number of asylum cases a judge faces in a given day, a proxy for the workload the judge faces (as caseload can have a significant impact on judicial decision-making, see Engel & Weinshall-Margel, 2020); and (d) the number of lenient decisions (grants) among the last five cases adjudicated by the same judge or in the same court, controlling for possible time variation in case

quality at the judge and court levels, respectively. The coefficient for serial position in Model 2 remains positive and significant ($\beta = 0.1012$, 95% CI [0.080, 0.122], $p < .001$, $OR = 1.107$).

Testing for the Effect of Breaks

In the current context, increasing leniency with serial position is difficult to explain under fatigue-based mechanisms that likely predict either increasing harshness (due to increase in choice of default/status quo) or no effect (due to effort differences). Here, we provide further evidence that fatigue is unlikely to explain our findings. If fatigue is a major driver of the serial position effect, then breaks, especially long ones, should eliminate the effect, and it should not carry over from 1 day to another. In contrast, Models 3 and 4 (Table 2) show that the rate of lenient decisions increases also as a function of the day of the week. Moreover, instead of analyzing the effect of serial position when it is defined within a sequence of daily cases, we can analyze the effect of serial position when it is defined within a sequence of weekly cases. Figure 1b shows that the rate of lenient decisions increases also as a function of the serial position of a case over the week, and Models 5 and 6 (Table 3) confirm the effect is statistically significant. Hence, even overnight breaks do not eliminate the serial position effect and fatigue (alone) cannot account for it.

Moderating Effect of Experience

We test for the moderating effect of experience on the serial position effect by including an interaction of serial position with the number of years of experience the judge has. Table 4 (and see Supplemental Table S4) shows the results of the interactions for sequences of daily (Model 7) and of weekly (Model 8) cases. In both models, the coefficient on the interaction between judges' experience and the serial position in the sequence is positive and significant. Figure 2 visualizes this moderating effect for representative values of judge experience (the first, second, and third quartiles). These results imply that consistent with motivational accounts, the apparent increasing leniency of decisions with serial position is stronger for more experienced judges.

Exploring the Different Motivational Accounts for the Serial Position Effect

As stated above, a quotas explanation for the serial position effect has two main predictions. First, it predicts negative autocorrelation between each two consecutive decisions. Indeed, in a previous analysis of these data, Chen et al. (2016) documented such negative autocorrelation. They explained it using a belief model akin to the gambler's fallacy. Importantly, they also tested for and rejected a quotas model as driving their findings. A second prediction of the quotas model is that controlling for the proportion of lenient decisions made so far in the sequence should eliminate the serial position effect. To examine this, we replaced the number of lenient decisions made by the judge out of the last five with the proportion of lenient decisions in the sequence so far (this necessarily discards the very first decision in a sequence). Supplemental Table S5 shows the robustness of the serial position effect to the addition of this control in both daily and weekly sequences. Together, the results

Table 3
Study 1 (Immigration Judges): Baseline Results, Weekly Sequences

	Lenient decision dummy	
	Model 5	Model 6
Variable	β (*SE*)	β (*SE*)
Serial position in the week	0.013*** (0.002)	0.038*** (0.003)
Republican judge		−0.089 (0.074)
Male judge		−0.413*** (0.078)
Age of judge		0.034*** (0.001)
Judge experience		
Over 8 years		−0.047** (0.015)
Government		−0.102 (0.074)
INS		−0.27*** (0.076)
Military		−0.379*** (0.111)
Private		−0.107 (0.076)
Academia		0.069 (0.147)
Lawyer present		1.160*** (0.022)
Defensive case		−0.502*** (0.009)
Applicant family size[a]		
2		0.253*** (0.015)
3		0.353*** (0.025)
4		0.504*** (0.032)
5		0.504*** (0.056)
6+		0.503*** (0.089)
Applicant nationality[b]		
China		0.091*** (0.012)
Haiti		−1.673*** (0.024)
El Salvador		−1.502*** (0.025)
Guatemala		−0.843*** (0.022)
Colombia		−0.482*** (0.022)
Number of cases in a week[a]		
2		−0.028 (0.015)
3		−0.068*** (0.016)
4		−0.079*** (0.017)
5		−0.095*** (0.019)
6		−0.163*** (0.022)
7		−−0.156*** (0.025)
8		−0.200*** (0.030)
9		−0.142*** (0.038)
10		−0.181*** (0.048)
11		−0.243*** (0.061)
12		−0.178 (0.103)
13		−0.152 (0.136)
14		0.039 (0.226)
15		−0.359 (0.411)
16		−0.463 (0.454)
Number of lenient decisions of last five in court[a]		
1		0.144*** (0.014)
2		0.219*** (0.014)
3		0.317*** (0.015)
4		0.470*** (0.018)
5		0.638*** (0.025)
Number of lenient decisions of last five of judge[a]		
1		0.250*** (0.014)
2		0.457*** (0.014)
3		0.658*** (0.016)
4		0.907*** (0.018)
5		1.244*** (0.023)
Constant	−1.13*** (0.116)	−3.181*** (0.141)
Random parts (*SD*s)		
Judge	0.870	0.638
Court	0.765	0.461
N	386,109	342,803

Note. Results of a mixed-effects logistic regression with random intercepts for judge and court. INS = Immigration and Naturalization Services; *SE* = standard error. Number of judges = 425 (Model 5) or 358 (Model 6). Number of courts = 53 (Model 5) or 52 (Model 6).
[a] Reference level for each of the ordinal variables is the minimal possible value (0 or 1). [b] Only the five most common nationalities were used, and the rest were classified as "other" (reference level).
** $p < .01$. *** $p < .001$.

Table 4
Study 1 (Immigration Judges): The Moderating Role of Experience

	Lenient decision dummy	
	Model 7	Model 8
Variable	β (*SE*)	β (*SE*)
Serial position in the day	0.048** (0.016)	
Serial position in the week		0.013** (0.005)
Judge experience	0.022*** (0.005)	0.023*** (0.005)
Serial Position in the Day × Judge Experience	0.006*** (0.001)	
Serial Position in the Week × Judge Experience		0.003*** (0.0004)
Constant	−1.943*** (0.231)	−1.855*** (0.231)
Random parts (*SD*s)		
Judge	0.600	0.600
Court	0.454	0.452
N	342,803	342,803

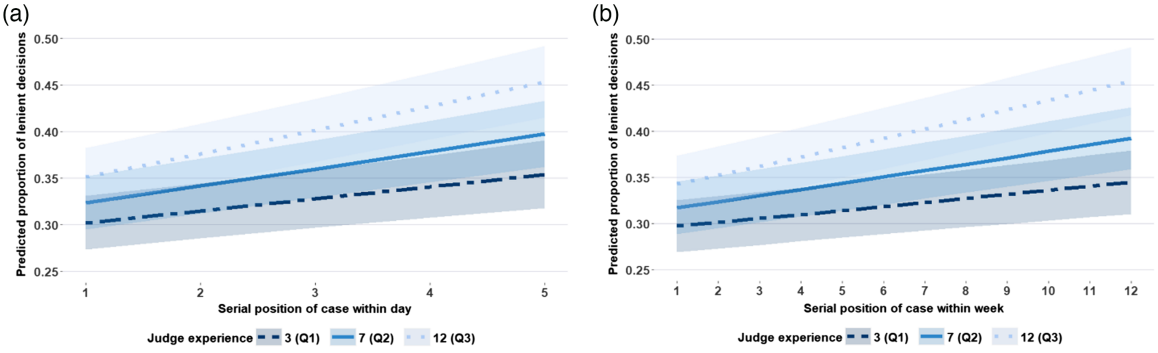
Note. *SE* = standard error. Results of a mixed-effects logistic regression with random intercepts for judge and court. Number of judges = 358. Number of courts = 52. Both models include control variables as in Models 2 and 6, respectively. The full models are provided in Supplemental Table S4 in the Supplemental Material.
** $p < .01$. *** $p < .001$.

suggest quotas are an unlikely explanation for the serial position effect we observe.

As discussed above, according to the calibration account, when anticipating additional future cases, judges will tend to avoid the extreme judgment, but this will not happen when the case is the only case of the day (Unkelbach & Memmert, 2014). Hence, under our assumption that lenient cases are the more extreme judgment, calibration predicts more lenient decisions for applications that were the only case in the sequence than in applications that happened to be positioned first in a multi-case sequence. The results do not support this prediction. The proportion of lenient decisions in days with a single case is 32.5%, but it is 35.9% for the first case in multiple-case days. However, there may be other differences between cases judged in single-case days and cases judged in multiple-case days (e.g., length of adjudication), and selection can therefore drive this difference. Hence, this result weighs against but does not rule out calibration as a mechanism driving the serial position effect.

According to the strategic consistency mechanism, people get increasingly likely to avoid decisions to come under increased scrutiny. Hence, in cases that are a priori less likely to be scrutinized, the serial position effect should be markedly smaller. In asylum hearings, defensive type cases are somewhat less likely to be appealed (vs. affirmative cases). Data from the BIA show that 51% of appeals are for defensive cases, although 58% of denials in our data are of defensive cases. Hence, if strategic consistency explains the serial position effect in immigration courts, we may expect a weaker effect for defensive type cases. Congruently with

Figure 2
Interaction of Serial Position and Judge Experience in Study 1



Note. Lines represent the predicted proportion of lenient decisions based on Models 7 (a) and 8 (b) from Table 4, for different serial positions when judge experience is held at its first, second, or third quartile, and with other predictors fixed (continuous predictors held constant at their mean and categorical predictors held at their proportional distribution in the data). Ribbons around lines represent 95% confidence intervals. See the online article for the color version of this figure.

this prediction, Supplemental Table S6 shows that, within a day, the serial position effect of defensive cases is significantly weaker than for affirmative cases. For weekly sequences, the serial position effects of defensive and affirmative cases do not significantly differ (but the point estimate is in the right direction), suggesting there may be additional contributors to the observed pattern. Nevertheless, the results of the daily sequences support a unique prediction of the strategic consistency account.

Robustness Checks

We performed a range of robustness checks, detailed in the Supplemental Material (and Supplemental Tables S7–S10), including a verification that the serial position effect we document is not driven by a systematic increase in quality of cases or by judges' heterogeneity as well as a replication of the results using fixed-effect regressions and for a subsample of days that include more than one case.

Discussion

We find that decisions by immigration judges get increasingly with the serial position of a case in a sequence of daily as well as of weakly asylum application hearings. The association between serial position and the probability of a lenient ruling is more pronounced the more experienced the judge is. Both the procedures in immigration courts for scheduling hearings to serial positions and our statistical tests suggest that sorting of cases by decision-relevant features is unlikely. Under the assumption that cases are indeed not sorted by (unobservable) increased quality, our results suggest that asylum decisions are influenced by a coincidental and immaterial feature of the asylum hearing: the later it happens to be scheduled in a judge's docket, the better the chances of the application to be granted, and this is particularly true when the application is heard by more experienced judges.

In the current context, the apparent increasing leniency of judges with serial position cannot be predicted by judges' fatigue that, as we discussed above, would predict either increasing harshness with serial position (if judges tend to preserve the status quo/default as implied by mental depletion) or no effect of serial position (if judges tend to choose less effortful options as implied by differences-in-effort mechanism). Fatigue explanations are also inconsistent with the observation that the effect persists over long breaks. In contrast, the leniency effect and the moderating role of judges' experience are both consistent with motivational accounts. In a follow-up exploratory analysis, we find that the pattern of results is mostly consistent with an increased tendency to avoid denials of applications that are more likely to be scrutinized (perhaps as a result of a desire to avoid having inconsistencies revealed) but cannot rule out an initial tendency to avoid approvals of asylum earlier in the sequence (perhaps as a result of a strategic way to preserve degrees of freedom in the early calibration stage). Regardless of the exact mechanism, however, judges' motivation to be, and perhaps more so to appear, consistent potentially drives the pattern of results.

Study 2: Jury Decisions in the Old Bailey 1751–1808

In Study 2, we analyze trial jury decisions made between 1751 and 1808 in the Old Bailey, a central criminal court in England. Here, we refer to acquittals as the lenient decision and to convictions as the harsh one. The most severe crimes (including all capital offenses) in two jurisdictions, City of London and County of Middlesex, were tried in the Old Bailey (Emsley et al., 2020). Trials were very short, rarely taking more than half an hour (Baker, 2019), with evidence suggesting the average trial length was 8 min (Feeley, 1997). After hearing each case, the juries usually went into huddles inside the courtroom and returned verdicts in a few minutes; they then continued immediately to hear the next case in the batch (Beattie, 1986). After hearing and deciding on all cases in the batch, the jury would retire to rest and return later to hear another batch of cases. When one jury panel rested, another jury panel, often of the other jurisdiction, would sit and decide on a different batch of cases. In our data, the median number of panels per session is 3 (range [1, 6]), and the mean number of cases per session is 74 ($SD = 30$).

We use data collected and processed by Bindler and Hjalmarsson (2019) and based on *The Proceedings of the Old Bailey*, a detailed account of criminal cases tried in the Old Bailey in those years. We follow an important assumption they make and validate with external sources: The order of cases as listed in the *Proceedings* reflects the true ordering of cases as presented to the judge and jury. We focus on juries in the Middlesex jurisdiction,[3] each hearing 30.7 trials on average ($SD = 12.7$). We exclude 5% of the observations that are missing identification for the judge and 0.3% of the observations in which the defendant pleaded guilty so the jury need not have given a verdict (but the serial position of trials accounts for these excluded observations). In total, we analyze 20,796 trials heard by 705 jury panels in 427 sessions. Supplemental Tables S11–S13 and the Supplemental Material provide further details on the data.

We first provide the predictions of each of the five mechanisms above for this setting and show they are likely similar to those presented for Study 1. Congruently, our analyses—and results—in this study are very similar to those of Study 1.

Derivation of Mechanism Predictions

Status Quo/Default (Increased Choice of the Status Quo/Default)

Defendants were usually held in jail from the time of their arrest until the end of their trial. Defendants were in a significant disadvantage as they normally could not know in advance the evidence presented against them nor could they enforce attendance of witnesses on their behalf. Although evidence of a defendant's guilt was required to support conviction, defendants were not presumed innocent until proven guilty. On the contrary, it was assumed that those who are truly innocent can disprove the (even circumstantial) evidence presented against them (Emsley et al., 2020). For example, absent defense counseling (which was uncommon), judges were those responsible to "give defendants every opportunity to prove their innocence," and

[3] Our analysis of the data shows that in the City of London jurisdiction, after 1760, capital offense cases tend to appear more frequently later in the sessions. Because capital offenses are more likely to be acquitted, increasing leniency as a function of serial position for the London jury is to be expected. Although we also control for the type of offense, making the observations conditionally independent, in the main analysis, we choose to focus on the jurisdiction in which we see no evidence for such sorting. Supplemental Figure S2 and Supplemental Table S14 include analyses for the London juries as well. Qualitatively, results are identical, although the point estimates of the effects for the London jurisdiction are somewhat larger.

defendants who refused to plead to the charges were deemed to be pleading guilty (Beattie, 1986). We take these aspects to suggest that the default verdict would have been, if anything, a conviction. Hence, increasing uptake of the default choice likely implies an increase in harsh decisions (convictions) as a function of serial position.

Differences in Effort (Increased Choice of the Less Effortful Decision)

While the available data do not allow computation of the decision time for different jury decisions, from what is known of the historical context, it seems that there should not be any difference in jury effort between lenient and harsh decisions. At the end of the hearing, when the judge charged the jury to return a verdict, they usually huddled inside the courtroom and returned a verdict within a few minutes. In so little time, the jury "could not review the evidence they had heard. There was neither time nor opportunity for one member of the jury to provide counter-arguments against what might appear immediately to be the majority view" (Beattie, 1986, p. 397). With such hasty discussion, there is no reason to believe one type of decision systematically required additional effort. Hence, as in Study 1, we assume this mechanism predicts no serial position effects in this context.

Quotas (Increased Choice of the Decision Restricted Under Quota)

In our data, 62% of jury decisions were harsh. In addition, similar arguments to those made above concerning the likely default decision suggest that if any decision is likely to be limited under quota it is the lenient one (acquitting). Hence, we assume this mechanism predicts that the jury will give increasingly more lenient decisions with serial position.

Calibration (Increased Choice of the More Extreme Decision)

As in Study 1, we assume that the decision that may be restricted under quota and the opposite of that which is considered the default (i.e., acquittal) is likely to also be considered the more extreme decision. Hence, an increase in extreme judgments due to calibration would predict an increase in lenient decisions as a function of serial position.

Strategic Consistency (Increased Choice of the Decision Likely to Be Scrutinized)

At the end of the sessions, defendants who were acquitted were released, whereas those who were convicted were brought in batches to be sentenced by the judge. Appeals on either jury decision were extremely rare. Hence, whereas lenient decisions meant the end of proceedings, harsh decisions were later scrutinized by at least one additional factor, the judge, who had considerable discretion over punishments in many cases. Some harsh decisions were scrutinized further. Specifically, all cases resulting with death sentences, which were mandatory for many offense types, had to be reviewed by the king's court (sometimes more than once), which led to pardons more often than not (Emsley et al., 2020). These procedures suggest that this mechanism predicts increasing leniency with serial position, and this prediction may be more pronounced for capital offense cases.

Admittedly, strategic consistency makes more sense for repeated actors hoping to preserve reputation than for mostly one-shot actors like jury panels (although nearly every panel consisted of experienced jurors, see below). Yet, the assumption that juries want to maintain consistency across cases is in line with previous conclusions drawn based on the same data (Bindler & Hjalmarsson, 2019). Furthermore, although jury panels had the sole final prerogative to reach a verdict, judges normally had considerable influence on them and their decisions.[4] Hence, the judges themselves may have had vested interests to avoid inconsistencies in verdicts that were later scrutinized by the king's court, and it is possible that their influence on the jury is one driving force of potential "strategic consistency" effects.

Analysis Strategy

Our analysis follows that of Study 1. It first establishes that, in line with the predictions of the motivational accounts but in contrast to those of the fatigue-based explanations, jury decisions get increasingly lenient with serial position. It then shows that this effect persists over breaks and is enhanced when the jury includes more experienced jurors. Finally, it explores some of the additional predictions of the motivational accounts and supports a prediction of the strategic consistency mechanism.

Because trials may include more than a single defendant (mean number of defendants per trial is 1.24, $SD = 0.59$), and because different defendants can get different verdicts, our main outcome variable is the share of defendants who received a lenient verdict (acquitted). To statistically test the effect of serial position, we implement linear mixed-effects regressions using R packages *lme4* (Bates et al., 2014) and *lmerTest* (Kuznetsova et al., 2017). We include random intercepts for jury panel, for batch of trials nested in jury panel, and for judge name and report restricted maximum likelihood estimates. Degrees of freedom were computed using Satterthwaite approximation. Data analysis also used packages *ggeffects* (Lüdecke, 2018) and *ggplot2* (Wickham, 2016).

Transparency and Openness

The data analyzed in this study were collected by Bindler and Hjalmarsson (2019). Our processing of the data is described in the Supplemental Material. The processed data and analysis code are available at https://osf.io/s8h2t/?view_only=c94d54fe57aa43fc84a8b6ccc350b9dc. The analysis plan was not preregistered. We adhered to the *Journal of Applied Psychology* methodological checklist.

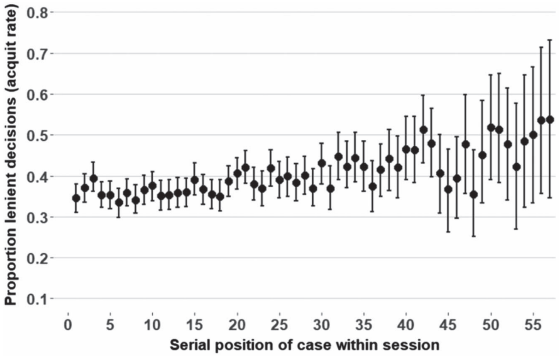
Results

Figure 3 plots the average proportion of lenient rulings as a function of the serial position of the trial within a jury session. Despite the fluctuations, it shows a clear increasing trend suggesting increasingly more lenient verdicts as a function of serial position.

[4] First, in the absence of legal counsel, judges played a major role in what, when, and how juries heard the evidence before them. Second, judges were very hands-on, often giving their own views on the case during and especially at the end of the trials. Indeed, our regressions below suggest the identity of the judge handling the case had a major impact on the verdict.

Figure 3

Proportion of Lenient Decisions (Acquittals) by the Serial Position of a Case in a Sequence of Cases in the Old Bailey



Note. Unadjusted descriptive statistics. Error bars indicate bootstrapped 95% confidence interval for the rates with 1,000 replicates. Cases in positions larger than 57 (145 cases, 0.6%) are omitted from the plot.

Table 5 shows the main results of the linear mixed-effects regressions. The main independent variable is the serial position of the trial within the session. Model 9 includes this variable as the only fixed effect. Its coefficient is positive and significant ($\beta = 0.0026$, 95% CI [0.0020, 0.0032], $p < .001$, $\eta^2 = 0.013$), suggesting verdicts become increasingly lenient with serial position. Model 10 controls for (a) observable case characteristics, namely the number of defendants on trial, the share of female defendants, and a detailed vector of 30 different offenses; (b) year fixed effects; and (c) proxies for load on the jury panel: the total number of trials in the session (across all panels in that session) and the total number of trials the focal jury panel faced in the session. The coefficient for serial position in the session remains positive and significant ($\beta = 0.0030$, 95% CI [0.0024, 0.0036], $p < .001$, $\eta^2 = 0.012$). Model 11 adds the proportion of acquittals in the last five cases decided by the same jury or in the same session (across juries), controlling for possible time variation in case quality at the panel and court levels, respectively. Notably, this excludes the first five decisions made by each jury. Again, the coefficient of serial position is positive and significant ($\beta = 0.0035$, 95% CI [0.0027, 0.0042], $p < .001$, $\eta^2 = 0.016$).

Testing the Effect of Breaks

If the observed pattern is triggered by increasing fatigue, breaks should have a major restorative impact on the prevalence of lenient verdicts. Recall that jury panels in the Old Bailey consecutively heard a batch of trials before taking a break (and sat through many such batches in each session). Unfortunately, breaks are not labeled in the *Proceedings*. To examine the effect of breaks, we define "minimal batches" as sequences of consecutive trials (trials with immediately consecutive IDs in the *Proceedings*) that are explicitly labeled to have been heard by the same jury and tried by the same judge. We assume that within minimal batches trials are unlikely to have been interrupted by breaks[5] (but actual batches are likely longer since, e.g., trials with missing jury or judge data cut the minimal batches short). Models 12–13 add to Models 10–11 the serial position within a minimal batch. The coefficient for serial position in the session is hardly affected, and models show no evidence for an effect within a minimal batch of cases.[6] This implies that the effect we observe depends on the total number of decisions made so far rather than on the number of decisions made since the last break.

Moderating Effect of Experience

None of the jurors in the Old Bailey was a professional juror. Yet, the same person often served in different juries over relatively short time spans. These more experienced jurors played a very dominant role in the jury decisions. According to Beattie (1986), "most often the majority must simply have acquiesced in a verdict arrived at by one or two dominant figures on the jury … it is here that previous experience on the jury was surely decisive" (pp. 397–398). In fact, experienced jurors often made decisions on behalf of the jury without consulting the other members (Beattie, 1986). If experienced jurors indeed played such a dominant role in the decision-making process, the serial position effect may be moderated by the juror's experience.

We define a juror's experience as the number of sessions over the previous 10 years in which a person with the same full name was part of a jury in the same jurisdiction (see Bindler & Hjalmarsson, 2019). We set the jury's experience to equal the experience of the most experienced juror. In our data, jury experience ranges between 0 and 10 with an *Mdn* of 3. To examine the moderating effect of jury experience, we add an interaction between the serial position of a trial in a session and jury experience. Models 14–15 (Table 6; and see Supplemental Table S15) show that the effect of increasing leniency with serial position is stronger for more experienced juries. Figure 4 visualizes this interaction for three levels of experience based on the quartiles of the distribution of jury experience.

Exploring the Different Motivational Accounts for the Serial Position Effect

The serial position effect we observe in this setting is unlikely driven by quotas. First, a quota model predicts negative correlation between each two consecutive decisions: all else equal, a current lenient decision decrements from the quota of lenient decisions and therefore increases the likelihood that the next decision would be harsh. Yet, as shown by Bindler and Hjalmarsson (2019), each two consecutive jury verdicts in the Old Bailey has strong *positive* correlation. Second, as with asylum decisions, we add to the

[5] The single courtroom in the Old Bailey was used by multiple judges and multiple jury panels during each session (97% of sessions in the data involve at least three judges and all sessions involve at least two panels). For a minimal batch to include a break, the courtroom should have been inactive between two consecutive trials with the same jury and judge. We follow Bindler and Hjalmarsson (2019) in assuming that this was unlikely as the nonactive jury (and judges) typically had to wait for their turn in rotation without the convenience of jury "withdrawing rooms."

[6] We also constructed a "maximal batch" variable: sequences of trials in the *proceedings* (not necessarily with consecutive IDs) that are tried by the same jury (see Supplemental Material). This variable captures sequences of trials for which there is no direct evidence that they were interrupted by a break. Replacing a minimal batch with a maximal batch does not change the results: no evidence for an effect within batches but a significant effect of serial position.

Table 5
Study 2 (Old Bailey): Baseline Results

	Share of lenient verdicts				
	Model 9	Model 10	Model 11	Model 12	Model 13
Variable	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)	β (*SE*)
Serial position in session	0.0026*** (0.0003)	0.0030*** (0.0003)	0.0035*** (0.0004)	0.0030*** (0.0003)	0.0035*** (0.0004)
Serial position in minimal batch				0.0002 (0.0021)	−0.0008 (0.0022)
Number of defendants		0.073*** (0.006)	0.068*** (0.007)	0.073*** (0.006)	0.068*** (0.007)
Share of female defendants		0.051*** (0.008)	0.052*** (0.008)	0.051*** (0.008)	0.052*** (0.008)
Offense					
Against the crown, major[a]		−0.084 (0.195)	−0.052 (0.216)	−0.084 (0.195)	−0.052 (0.216)
Against the crown, minor[a]		0.019 (0.068)	−0.030 (0.074)	0.019 (0.068)	−0.03 (0.074)
Animal theft[a]		0.018 (0.037)	0.027 (0.041)	0.018 (0.037)	0.028 (0.041)
Arson[a]		0.350** (0.117)	0.406** (0.139)	0.351** (0.117)	0.406** (0.139)
Assault		0.482 (0.325)	0.451 (0.327)	0.481 (0.325)	0.454 (0.327)
Bigamy		0.032 (0.064)	0.060 (0.068)	0.032 (0.064)	0.060 (0.068)
Burglary[a]		−0.009 (0.035)	0.006 (0.038)	−0.009 (0.035)	0.006 (0.038)
Coining[a]		−0.026 (0.046)	−0.041 (0.050)	−0.026 (0.046)	−0.041 (0.050)
Embezzlement		0.229*** (0.061)	0.257*** (0.065)	0.229*** (0.061)	0.257*** (0.065)
Forgery[a]		0.203*** (0.046)	0.227*** (0.051)	0.204*** (0.046)	0.227*** (0.051)
Fraud[a]		−0.059 (0.097)	−0.141 (0.107)	−0.059 (0.097)	−0.141 (0.107)
House-breaking[a]		−0.128** (0.044)	−0.138** (0.050)	−0.127** (0.044)	−0.138** (0.050)
Infanticide[a]		0.463*** (0.077)	0.472*** (0.083)	0.463*** (0.077)	0.472*** (0.083)
Larceny		−0.028 (0.033)	−0.019 (0.036)	−0.028 (0.033)	−0.019 (0.036)
Mail theft[a]		0.258 (0.210)	0.289 (0.210)	0.259 (0.210)	0.289 (0.210)
Manslaughter		0.413*** (0.114)	0.397*** (0.118)	0.413*** (0.114)	0.397*** (0.118)
Murder[a]		0.091* (0.042)	0.100* (0.046)	0.091* (0.042)	0.100* (0.046)
Perjury		0.634** (0.210)	0.622** (0.210)	0.635** (0.21)	0.621** (0.21)
Perverting-justice[a]		−0.021 (0.081)	−0.035 (0.088)	−0.021 (0.081)	−0.035 (0.088)
Pickpocketing[a]		0.124** (0.038)	0.143*** (0.042)	0.123** (0.038)	0.143*** (0.042)
Rape[a]		0.462*** (0.057)	0.480*** (0.063)	0.462*** (0.057)	0.480*** (0.063)
Receiving stolen goods[a]		0.139*** (0.037)	0.148*** (0.040)	0.139*** (0.037)	0.148*** (0.040)
Returning transportation[a]		−0.105 (0.055)	−0.101 (0.061)	−0.105 (0.055)	−0.101 (0.061)
Riot[a]		0.219* (0.087)	0.250** (0.094)	0.219* (0.087)	0.250** (0.094)
Robbery[a]		0.064 (0.035)	0.089* (0.038)	0.064 (0.035)	0.089* (0.038)
Shoplifting[a]		−0.144*** (0.038)	−0.133** (0.042)	−0.144*** (0.038)	−0.133** (0.042)
Sodomy[a]		0.240* (0.097)	0.176 (0.115)	0.240* (0.097)	0.176 (0.115)
Theft-from-place[a]		−0.007 (0.034)	0.006 (0.037)	−0.007 (0.034)	0.006 (0.037)
Wounding		0.303*** (0.068)	0.315*** (0.080)	0.303*** (0.068)	0.315*** (0.080)
Average share of acquittals in last five trials by jury			−0.038 (0.021)		−0.038 (0.021)
Average share of acquittals in last five trials in session			0.127*** (0.021)		0.127*** (0.021)
constant	0.335*** (0.009)	0.524 (0.512)	−0.051 (0.520)	0.523 (0.513)	−0.050 (0.520)
Random parts (*SD*s)					
Minimal batch: Jury	0.115	0.116	0.096	0.116	0.096
Jury	0.044	0.027	0.029	0.027	0.029
Judge	0.042	0.035	0.035	0.035	0.035
σ	0.454	0.445	0.450	0.445	0.450
N	20,796	20,783	17,002	20,783	17,002

Note. *SE* = standard error. Results of a mixed-effects linear regression with random intercepts for judge, jury, and "minimal batch" (see Supplemental Material) nested within jury. Number of minimal batches within jury = 13,319 (Model 9), 13,310 (Models 10, 12), or 10,678 (Models 11, 13). Number of juries = 705 (Models 9, 10, 12) or 698 (Models 11, 13). Number of judges = 67. Reference level for offense type is "other." Models 10–13 also include fixed effects for year (54 dummy variables), total number of session cases (116 dummies), and total number of jury cases (68 in Models 10, 12; 63 in Models 11, 13). Coefficients of these are not shown and can be retrieved from the authors.
[a] Capital offense.
* $p < .05$. ** $p < .01$. *** $p < .001$.

baseline models the proportion of acquittals the jury made in the session so far. Supplemental Table S16 shows this does not influence the effect of serial position.

The serial position effect is also somewhat less likely to be driven by calibration. In the main analyses, verdicts are treated as binary (acquit or convict), yet in practice, it was possible to convict defendants with reduced, rather than original, charges. Jury often convicted with reduced charges to avoid convictions with felonies that mandated capital punishment. Hence, convictions with reduced charges may be a middle ground between the two more extreme categories. Because calibration predicts an increase in extreme judgments with serial position, it should predict not only an increase in the most lenient verdict (acquittals) but also an increase in the harshest verdict (convictions with original charges). This prediction,

Table 6
Study 2 (Old Bailey): The Moderating Role of Experience

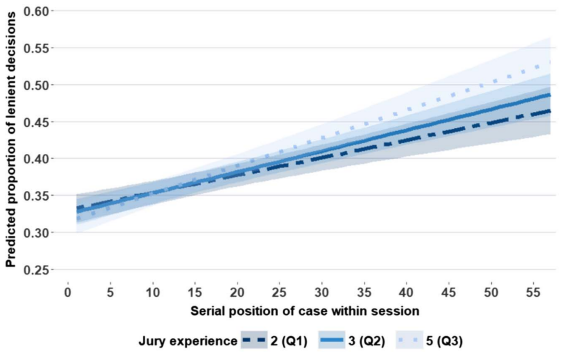
	Share of lenient verdicts	
	Model 14	Model 15
Variable	β (*SE*)	β (*SE*)
Serial position in session	0.0014* (0.0006)	0.0019** (0.0007)
Jury experience	−0.005 (0.003)	−0.006 (0.004)
Serial Position in Session × Jury Experience	0.0005*** (0.0001)	0.0005** (0.0002)
Constant	0.558 (0.514)	0.158 (0.334)
Random parts (*SD*s)		
Minimal batch: Jury	0.109	0.090
Jury	0.026	0.023
Judge	0.036	0.036
σ	0.447	0.451
N	20,249	16,575

Note. *SE* = standard error. Results of a mixed-effects linear regression with random intercepts for judge, jury, and "minimal batch" (see Supplemental Material) nested within jury. Number of minimal batches within jury = 12,949 (Model 14) or 10,384 (Model 15). Number of juries = 679 (Model 14) or 673 (Model 15). Number of judges = 65. Models include all control variables as in Models 10–11 from Table 5. Supplemental Table S15 (Supplemental Material) provides details for the full models.
* $p < .05$. ** $p < .01$. *** $p < .001$.

however, is not supported by the data (Supplemental Table S17): As with all convictions pooled, the rate of convictions with original charges decreases as a function of serial position. Nevertheless, the qualitative difference between acquittal and conviction with any charge is intuitively much greater than the difference between the two types of convictions, hence this finding does not completely rule out calibration as an explanation for the serial position effect.

In contrast to the other two motivational accounts, the data support one prediction of the strategic consistency account. According to strategic consistency, the serial position effect is driven by an increasing tendency to avoid decisions that come under scrutiny. In the current context, convictions on capital offense cases come under increased scrutiny. Hence, the strategic consistency explanation predicts a stronger serial position effect in capital than in noncapital offense cases. The data support this unique prediction of the strategic consistency account (Supplemental Table S18).

Figure 4
Interaction of Serial Position and Jury Experience in Study 2



Note. Lines represent the predicted proportion of lenient decisions based on Model 14 (Table 6) for different serial positions when jury experience is held at its first, second, and third quartile and with other predictors fixed (continuous predictors held constant at their mean and categorical predictors held at their proportional distribution in the data). Ribbons around lines represent 95% confidence intervals. Cases in positions larger than 57 (145 cases, 0.6%) are omitted from the plot. See the online article for the color version of this figure.

Robustness Checks

Although the historical context makes it hard to ascertain that there are not any unobservable characteristics that drive the serial position effect we observe, a range of robustness checks, detailed in the Supplemental Material (and Supplemental Tables S19–S23), suggests that this is unlikely.

Discussion

The results of Study 2 show that jury decisions in 18th century England criminal court got increasingly more lenient as a function of serial position of the case in a jury session. Moreover, this serial position effect is stronger among juries with more experienced jurors and is not impacted by the timing of the most recent break. These results are predicted by motivational accounts of serial position effects, but not by fatigue accounts, lending further support to the idea that decision-makers' motivations for consistency may drive serial position effects in important high-stakes settings. Furthermore, a unique prediction of one of the motivational mechanisms, strategic consistency, is confirmed in the current data. Specifically, the increasing leniency with serial position is stronger in cases for which harsh verdicts will surely be scrutinized to a greater extent.

The context in 18th century Old Bailey is clearly different from that of the legal system today, and the structure of the sequence of decisions that juries in the Old Bailey faced is very different from the structure of the sequence of decisions faced by immigration judges today (Study 1). For example, immigration judges face sequences of very few decisions made after hours-long hearings, whereas juries in the Old Bailey faced sequences of many decisions made after minutes-long trials. Nevertheless, the results in Study 2 echo all the main conclusions of Study 1, including a leniency effect of serial

position and the moderating role of experience and breaks. In both studies, the results seem most in line with a strategic consistency explanation. The consistency of results between the two studies is compelling considering the large differences between the contexts and enhances confidence that the main conclusions are generalizable, at least in legal contexts.

General Discussion

Our analysis of two legal contexts with very high stakes reveals decisions got increasingly lenient as a function of their serial position in a sequence. Our theory and data support the notion that this increasing leniency with serial position is driven, at least in part, by decision-makers' motivation to be or appear consistent. Specifically, in these contexts, three theoretical mechanisms based on the motivation for consistency predict increasing leniency, whereas more common explanations for serial position effects based on fatigue predict the opposite. Additional predictions of the motivational mechanisms—a stronger effect of more experienced decision-makers and a stronger effect in decisions more likely to come under scrutiny—are also supported by the data. These results have important implications for the study of serial position effects and for many domains in which decision-makers need to make independent sequential decisions.

Theoretical Implications and Future Research

Serial position effects, a systematic impact of the serial position of a case in sequence on a decision, have been documented in many contexts (Augenblick & Nicholson, 2016; Bian et al., 2022; Colton & Peterson, 1967; Danziger et al., 2011; Hirshleifer et al., 2019; Ibanez & Toffel, 2020; Linder et al., 2014; McLaughlin et al., 2009; Memmert et al., 2008; Orazbayev, 2017; Philpot et al., 2018; Unkelbach et al., 2012). To explain them, previous research focused on the role of decision-makers' fatigue. We document serial position effects in important field settings that are likely driven by a very different reason, decision-makers' motivated behavior. Our investigation parallels recent studies showing that behavioral phenomena commonly explained by cognitive factors can often be explained by people's motivations (Dorison et al., 2022; Dorison & Heller, 2022; Mobius et al., 2022; Van Zant, 2022), clarifying why these phenomena are common and when they can be expected.

Specifically, our results highlight a paradox: Systematic inconsistencies in treatment of different cases in sequence may be particularly common when decision-makers have high motivation to behave consistently (or at least appear as if they do). Because consistency in decisions signals competence (Falk & Zimmermann, 2017) and is considered a necessary condition for expertise (Einhorn, 1974), organizational contexts in which these systematic inconsistencies emerge may be more common than currently thought. Moreover, as our theoretical analysis highlights, and as our data shows, the most experienced decision-makers are those most likely to exhibit this bias.

Our theoretical analysis suggests three distinct channels through which motivation to be or appear consistent leads to serial position effects. Among the three, in our empirical contexts, the data are most consistent with the strategic consistency mechanism that predicts an increase in choice of the option that comes under less scrutiny. One implication of this mechanism, supported by the data, is that the serial position effect will be stronger in cases where decision-makers can expect the most scrutiny on their decisions.

Although our data are less consistent with the quotas and calibration mechanisms, in other contexts, any of the three channels may play a larger role. For example, past studies have already demonstrated the importance of calibration effects (Unkelbach & Memmert, 2014). We do not feel the current investigation, which focused on motivation for consistency in general (contrasted with fatigue), allows sufficient distinction between the three channels. To study the relative importance of these distinct channels, a rigorous theoretical investigation that derives unique predictions of each of them is in order.

One possible direction that would allow conducting a more careful comparison of the three mechanisms in a constructive manner is to rely on more basic research on sequential categorical judgments. In a typical experiment, participants are presented with a stream of stimuli (e.g., squares of varying sizes) and are asked to categorize them (e.g., to "large" or "small"). While much more abstract than the decisions we deal with, it is possible to draw links between several theoretical principles suggested in that line of research and the motivational accounts we propose. For example, akin to the logic underlying the quotas mechanism, it has been suggested that participants expect to use the possible categories with certain frequencies and that these expectations change how they categorize the stimuli (Stevens & Galanter, 1957). Similarly, it has been suggested that judgments are made in relation to the most extreme stimuli in the available context (Parducci, 1965), one of the core ideas underlying the calibration mechanism. Finally, a desire for consistency has been suggested as a major source for many of the phenomena that emerge in sequential categorical judgments (Haubensak, 1992). Alone, none of these principles is sufficient to account for serial position effects as further assumptions are necessary. Yet, there are clear parallels between these basic studies and the potential sources of the more complex phenomena. Hence, it should be possible to design simple experiments using abstract stimuli that help disentangle the contribution of each motivation-based mechanism to serial position effects.

An additional question left for future research concerns the impact of changes in the environment and of time on the effects of the motivation to be or appear consistent. Consistency is likely desirable only to the extent that circumstances have not changed. If they have, consistency may imply stubbornness and rigidity. Therefore, the influence of the motivation to be consistent on behavior likely fades as the environment changes and/or as a function of time. In our analysis of Study 1, we find an effect consistent with motivational explanations within a day and within a week, and the latter seems notably smaller than the former. Future studies should, however, directly test the prediction that the effects fade with time or when circumstances clearly change.

It is important to note that while in the current contexts, serial position effects are more likely driven by motivation than by fatigue, it does not imply that fatigue is not an important driver of serial position effects in other contexts. In our empirical contexts, no one decision systematically elicits more time and effort than another, and the differences-in-effort mechanism that implies increased choice of the less effortful option has no clear predictions. Hence, our results do not challenge the notion that fatigue leads to serial position effects when there *is* an asymmetry in time or effort required by different decisions. Indeed, while our study cannot be used to

support it, we expect that in contexts in which such an asymmetry exists, fatigue will play a more significant role in driving serial position effects.

In contrast, the predictions of a second channel through which fatigue may potentially lead to serial position effects, increased choice of the default/status quo option as a result of mental depletion, are directly rejected by our results. Furthermore, to the best of our knowledge, there is no clear evidence for the status quo/default mechanism as a driver of serial position effects in contexts in which the effect cannot be attributable to the more parsimonious "differences-in-effort" mechanism. For example, the pattern of increased harshness in Israeli parole hearings (Danziger et al., 2011) can be explained as an increase in choice of the status quo (as suggested by the authors), but it is also consistent with an increase in choice of the alternative that elicits less time and effort. Combined with our findings, we feel this observation challenges the very existence of an independent status quo/default mechanism. Future studies should directly test the claim that, after making multiple decisions, individuals tend to choose the default—rather than the less effortful—option more.

Although we feel our study can be used to reject the status quo/default mechanism, the fact that in our contexts a major fatigue mechanism, differences-in-effort, has no clear predictions raises a wider question concerning the relative importance and contribution of fatigue-based versus motivational drivers of serial position effects, when both are likely present. Future research should try to evaluate settings in which fatigue mechanisms have stronger directional predictions that directly contrast those of the motivational accounts. When the two classes predict opposite effects, do they cancel each other, or is one of them stronger than the other? One interesting hypothesis is that in such cases, more experienced decision-makers would behave in accordance with the predictions of the motivational accounts and less experienced decisions-makers would behave in accordance to the predictions of the fatigue explanation.

Practical Implications

In our investigation of two very different legal contexts, we find a very similar pattern of results and specifically increasing leniency with serial position. To the extent this pattern is generalizable to decision-makers in other similar contexts (judges in other courts, and also auditors; inspectors of quality, safety, and environmental standards; and moderators of online content or violation reports), a clear implication of the current research is that individuals who happen to appear later in a decision-maker's docket are likelier to get more lenient treatment. The fact that our results are most in line with the strategic consistency account implies that this is particularly true in contexts in which decision scrutiny and effective appeal are practical. Because often the motivation of an individual to appeal a harsh decision is much greater than the motivation of an institution to appeal a lenient one, in such contexts, harsh decisions would likely be scrutinized more. This also means that to reduce the impact of strategic consistency considerations, officials may need to scrutinize lenient decisions to a greater extent.

Although increasing leniency with serial position may be a common pattern in contexts most similar to those we study in this article, it is not necessarily the most likely pattern in general. To predict whether and to what direction serial position effects emerge in their own organizations, practitioners can use the taxonomy of mechanisms for serial position effects that we have presented. The existence and direction of a serial position effect in a particular context will likely depend on the implications of a certain decision relative to another. Does it imply much more work and effort for the decision-maker? Does it limit the decision-maker's degrees of freedom when making additional decisions? Is it much more likely to come under scrutiny and if so, are previous decisions in the sequence transparent? When the answer to one of these questions is positive, stakeholders should know that it may be avoided in later segments of sequences of decisions.

Identifying the mechanism (or mechanisms) responsible for a particular serial position effect is key to development of interventions that would reduce it and its adverse consequences. Specifically, when the effect is triggered mainly by decision-maker fatigue, reducing decision load and the addition of breaks between consecutive decisions should be helpful, but these interventions are unlikely to be useful when the effect is triggered by motivation for consistency.

In those cases, allowing decision-makers to review all cases in the sequence before they have to deliver any decisions may be helpful, as this would allow them to then deliver decisions that are already consistent with one another (or with some norm) without having to bias some decisions in the sequence (see a related intervention in Fasold et al., 2015). Yet, in many contexts, such change of decision procedures would be impractical. For example, when decisions must be delivered immediately after the case is presented (like it was in the Old Bailey), or when workload on decision-makers is already high and adding another "round of review" for each sequence is infeasible (like it is in immigration courts). Moreover, with such a change of procedures, different types of biases might emerge (e.g., due to regression to the mean in memory recall, evaluations may become less extreme with time; see Li & Epley, 2009).

Another potential way to reduce motivation-based serial position effects is to hold decision-makers who exhibit them accountable. Specifically, motivational mechanisms assume decision-makers wish to avoid appearing biased or incompetent; if they know that their sequences of decisions can be directly checked for serial position effects, this may produce increased effort to avoid them. Again, however, this may be impractical in many settings, for example, when sequences of decisions are not made very often, and it is hard to find evidence that the observed pattern is unjustified.

When direct interventions are impractical, randomization of the order of cases that appear in a sequence may increase fairness to some degree. While randomization alone will not eliminate serial position effects, at least all involved parties can have equal chance to be assigned to "favorable" positions. Yet, scheduling often depends on the parties' (reported) availability. Full randomization would then be infeasible, and to make things worse, parties may attempt to strategically influence their serial positions to get more favorable outcomes. To maintain fairness and bias-free decisions, it is therefore essential to find effective interventions that would reduce these effects.

Conclusion

Many settings involve making supposedly independent decisions in sequence. We document, in contexts in which fairness is crucial and stakes cannot be higher, how such decisions are influenced by

their mere position in the sequence. The fact that the decisions are made by highly trained professional decision-makers and the bias is even stronger among the more experienced of them emphasizes the importance of identifying the roots of the phenomena. Our work makes a step in this direction by highlighting that decision-makers' motivation to be or appear consistent across decisions and time can be a major driver of serial position effects.

References

Antipov, E. A., & Pokryshevskaya, E. B. (2017). Order effects in the results of song contests: Evidence from the eurovision and the new wave. *Judgment and Decision Making*, *12*(4), 415–419. https://journal.sjdm.org/16/16121/jdm16121.pdf

Ash, E., Chen, D. L., Andrus, C., Godevais, D., & Ng, G. (2020). *Machine prediction of appeal success in U.S. Asylum Courts*. https://www.nber.org/~dlchen/papers/Machine_Prediction_of_Appeal_Success_in_US_Asylum_Courts.pdf

Assy, R., & Menashe, D. (2014). The catch-22 in Israel's parole law. *Criminal Justice and Behavior*, *41*(12), 1422–1436. https://doi.org/10.1177/0093854814548448

Augenblick, N., & Nicholson, S. (2016). Ballot position, choice fatigue, and voter behaviour. *The Review of Economic Studies*, *83*(2), 460–480. https://doi.org/10.1093/restud/rdv047

Baker, J. (2019). *Introduction to English legal history* (5th ed.). Oxford University Press. https://doi.org/10.1093/oso/9780198812609.001.0001

Bartlett, F. C. (1943). Ferrier Lecture-Fatigue following highly skilled work. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *131*(864), 247–257. https://doi.org/10.1098/rspb.1943.0006

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4*. ArXiv Preprint ArXiv:1406.5823. https://doi.org/10.48550/arXiv.1406.5823

Beattie, J. M. (1986). *Crime and the courts in England, 1660–1800* (Vol. 404). Princeton University Press Princeton.

Bhargava, S., & Fisman, R. (2014). Contrast effects in sequential decisions: Evidence from speed dating. *The Review of Economics and Statistics*, *96*(3), 444–457. https://doi.org/10.1162/REST_a_00416

Bian, J., Greenberg, J., Li, J., & Wang, Y. (2022). Good to go first? Position order effects in expert evaluation of early-stage ventures. *Management Science*, *68*(1), 300–315. https://doi.org/10.1287/mnsc.2021.4132

Bindler, A., & Hjalmarsson, R. (2019). Path dependency in jury decision making. *Journal of the European Economic Association*, *17*(6), 1971–2017. https://doi.org/10.1093/jeea/jvy046

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2005). Effects of mental fatigue on attention: An ERP study. *Cognitive Brain Research*, *25*(1), 107–116. https://doi.org/10.1016/j.cogbrainres.2005.04.011

Bruine de Bruin, W. (2005). Save the last dance for me: Unwanted serial position effects in jury evaluations. *Acta Psychologica*, *118*(3), 245–260. https://doi.org/10.1016/j.actpsy.2004.08.005

Bruine de Bruin, W. (2006). Save the last dance II: Unwanted serial position effects in figure skating judgments. *Acta Psychologica*, *123*(3), 299–311. https://doi.org/10.1016/j.actpsy.2006.01.009

Chen, D. L., Moskowitz, T. J., & Shue, K. (2016). Decision making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires. *The Quarterly Journal of Economics*, *131*(3), 1181–1242. https://doi.org/10.1093/qje/qjw017

Cohen, D., Plonsky, O., & Erev, I. (2020). On the impact of experience on probability weighting in decisions under risk. *Decision*, *7*(2), 153–162. https://doi.org/10.1037/dec0000118

Colton, T., & Peterson, O. L. (1967). An assay of medical students' abilities by oral examination. *Journal of Medical Education*, *42*(11), 1005–1014.

Dai, H., Milkman, K. L., Hofmann, D. A., & Staats, B. R. (2015). The impact of time at work and time off from work on rule compliance: The case of hand hygiene in health care. *Journal of Applied Psychology*, *100*(3), 846–862. https://doi.org/10.1037/a0038067

Daljord, Ø., Urminsky, O., & Ureta, J.-M. (2019). *A test of the status quo theory of depletion: Evidence from the Israeli parole decisions*. https://pdfs.semanticscholar.org/9f3b/28022a4cea1d42d1caab006526f046805fc7.pdf

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(17), 6889–6892. https://doi.org/10.1073/pnas.1018033108

Deng, H., Coyle-Shapiro, J., & Yang, Q. (2018). Beyond reciprocity: A conservation of resources view on the effects of psychological contract violation on third parties. *Journal of Applied Psychology*, *103*(5), 561–577. https://doi.org/10.1037/apl0000272

Dhami, M. K. (2003). Psychological models of professional decision making. *Psychological Science*, *14*(2), 175–180. https://doi.org/10.1111/1467-9280.01438

Dorison, C. A., & Heller, B. H. (2022). Observers penalize decision makers whose risk preferences are unaffected by loss-gain framing. *Journal of Experimental Psychology: General*, *151*(9), 2043–2059. https://doi.org/10.1037/xge0001187

Dorison, C. A., Umphres, C. K., & Lerner, J. S. (2022). Staying the course: Decision makers who escalate commitment are trusted and trustworthy. *Journal of Experimental Psychology: General*, *151*(4), 960–965. https://doi.org/10.1037/xge0001101

Einhorn, H. J. (1974). Expert judgment: Some necessary conditions and an example. *Journal of Applied Psychology*, *59*(5), 562–571. https://doi.org/10.1037/h0037164

Emsley, C., Hitchcock, T., & Shoemaker, R. (2020). *Crime and justice—Trial procedures*. Old Bailey Proceedings Online. www.oldbaileyonline.org

Engel, C., & Weinshall-Margel, K. (2020). Manna from heaven for judges–judges' reaction to a quasi-random reduction in caseload. *Journal of Empirical Legal Studies*, *17*(4), 722–751. https://doi.org/10.1111/jels.12265

Falk, A., & Zimmermann, F. (2017). Consistency as a signal of skills. *Management Science*, *63*(7), 2197–2210. https://doi.org/10.1287/mnsc.2016.2459

Fasold, F., Memmert, D., & Unkelbach, C. (2012). Extreme judgments depend on the expectation of following judgments: A calibration analysis. *Psychology of Sport and Exercise*, *13*(2), 197–200. https://doi.org/10.1016/j.psychsport.2011.11.004

Fasold, F., Memmert, D., & Unkelbach, C. (2015). A theory-based intervention to prevent calibration effects in serial sport performance evaluations. *Psychology of Sport and Exercise*, *18*, 47–52. https://doi.org/10.1016/j.psychsport.2015.01.001

Feeley, M. M. (1997). Legal complexity and the transformation of the criminal process: The origins of plea bargaining. *Israel Law Review*, *31*(1–3), 183–222. https://doi.org/10.1017/S0021223700015284

Finkelstein, S., & Hambrick, D. C. (1990). Top-management-team tenure and organizational outcomes: The moderating role of managerial discretion. *Administrative Science Quarterly*, *35*(3), 484–503. https://doi.org/10.2307/2393314

Glejser, H., & Heyndels, B. (2001). Efficiency and inefficiency in the ranking in competitions: The case of the Queen Elisabeth music contest. *Journal of Cultural Economics*, *25*(2), 109–129. https://doi.org/10.1023/A:1007659804416

Glöckner, A. (2016). The irrational hungry judge effect revisited: Simulations reveal that the magnitude of the effect is overestimated. *Judgment and Decision Making*, *11*(6), 601–610.

Haan, M. A., Dijkstra, S. G., & Dijkstra, P. T. (2005). Expert judgment versus public opinion–evidence from the Eurovision song contest. *Journal of Cultural Economics*, *29*(1), 59–78. https://doi.org/10.1007/s10824-005-6830-0

Haerem, T., & Rau, D. (2007). The influence of degree of expertise and objective task complexity on perceived task complexity and performance. *Journal of Applied Psychology*, *92*(5), 1320–1331. https://doi.org/10.1037/0021-9010.92.5.1320

Hartzmark, S. M., & Shue, K. (2018). A tough act to follow: Contrast effects in financial markets. *The Journal of Finance*, *73*(4), 1567–1613. https://doi.org/10.1111/jofi.12685

Haubensak, G. (1992). The consistency model: A process model for absolute judgments. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 303–309. https://doi.org/10.1037/0096-1523.18.1.303

Hirshleifer, D., Levi, Y., Lourie, B., & Teoh, S. H. (2019). Decision fatigue and heuristic analyst forecasts. *Journal of Financial Economics*, *133*(1), 83–98. https://doi.org/10.1016/j.jfineco.2019.01.005

Hockey, G. R. J. (2013). *The psychology of fatigue: Work, effort and control*. Cambridge University Press. https://doi.org/10.1017/CBO9781139015394

Hockey, G. R. J., & Earle, F. (2006). Control over the scheduling of simulated office work reduces the impact of workload on mental fatigue and task performance. *Journal of Experimental Psychology: Applied*, *12*(1), 50–65. https://doi.org/10.1037/1076-898X.12.1.50

Hsiang, E. Y., Mehta, S. J., Small, D. S., Rareshide, C. A. L., Snider, C. K., Day, S. C., & Patel, M. S. (2019). Association of primary care clinic appointment time with clinician ordering and patient completion of breast and colorectal cancer screening. *JAMA Network Open*, *2*(5), Article e193403. https://doi.org/10.1001/jamanetworkopen.2019.3403

Hull, C. L. (1943). *Principles of behavior*. Appleton-Century.

Hunter, E. M., & Wu, C. (2016). Give me a better break: Choosing workday break activities to maximize resource recovery. *Journal of Applied Psychology*, *101*(2), 302–311. https://doi.org/10.1037/apl0000045

Ibanez, M. R., & Toffel, M. W. (2020). How scheduling can bias quality assessment: Evidence from food-safety inspections. *Management Science*, *66*(6), 2291–2799. https://doi.org/10.1287/mnsc.2019.3318

Kim, B., Park, S., & Zhao, Y. (2021). How people vote in contests: New findings from Immortal Songs 2. *Journal of Cultural Economics*, *45*(1), 45–62. https://doi.org/10.1007/s10824-020-09386-0

Kim, R. H., Day, S. C., Small, D. S., Snider, C. K., Rareshide, C. A. L., & Patel, M. S. (2018). Variations in influenza vaccination by clinic appointment time and an active choice intervention in the electronic health record to increase influenza vaccination. *JAMA Network Open*, *1*(5), Article e181770. https://doi.org/10.1001/jamanetworkopen.2018.1770

Kim, S., Park, Y., & Headrick, L. (2018). Daily micro-breaks and job performance: General work engagement as a cross-level moderator. *Journal of Applied Psychology*, *103*(7), 772–786. https://doi.org/10.1037/apl0000308

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Leibovitch, A. (2016). Relative judgments. *The Journal of Legal Studies*, *45*(2), 281–330. https://doi.org/10.1086/687376

Levav, J., Heitmann, M., Herrmann, A., & Iyengar, S. S. (2010). Order in product customization decisions: Evidence from field experiments. *Journal of Political Economy*, *118*(2), 274–299. https://doi.org/10.1086/652463

Li, Y. E., & Epley, N. (2009). When the best appears to be saved for last: Serial position effects on choice. *Journal of Behavioral Decision Making*, *22*(4), 378–389. https://doi.org/10.1002/bdm.638

Linder, J. A., Doctor, J. N., Friedberg, M. W., Reyes Nieva, H., Birks, C., Meeker, D., & Fox, C. R. (2014). Time of day and the decision to prescribe antibiotics. *JAMA Internal Medicine*, *174*(12), 2029–2031. https://doi.org/10.1001/jamainternmed.2014.5225

Lorist, M. M., Boksem, M. A. S., & Ridderinkhof, K. R. (2005). Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cognitive Brain Research*, *24*(2), 199–205. https://doi.org/10.1016/j.cogbrainres.2005.01.018

Lüdecke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, *3*(26), Article 772. https://doi.org/10.21105/joss.00772

Lyons, M., Al-Nakeeb, Y., & Nevill, A. (2006). The impact of moderate and high intensity total body fatigue on passing accuracy in expert and novice basketball players. *Journal of Sports Science & Medicine*, *5*(2), 215–227.

Mantonakis, A., Rodero, P., Lesschaeve, I., & Hastie, R. (2009). Order in choice: Effects of serial position on preferences. *Psychological Science*, *20*(11), 1309–1312. https://doi.org/10.1111/j.1467-9280.2009.02453.x

McLaughlin, K., Ainslie, M., Coderre, S., Wright, B., & Violato, C. (2009). The effect of differential rater function over time (DRIFT) on objective structured clinical examination ratings. *Medical Education*, *43*(10), 989–992. https://doi.org/10.1111/j.1365-2923.2009.03438.x

Memmert, D., Unkelbach, C., Ertmer, J., & Rechner, M. (2008). Gelb oder kein Gelb? [Yellow or no yellow?]. *Zeitschrift für Sportpsychologie*, *15*(1), 1–11. https://doi.org/10.1026/1612-5010.15.1.1

Mobius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2022). Managing self-confidence: Theory and experimental evidence. *Management Science*. Advance online publication. https://doi.org/10.1287/mnsc.2021.4294

Orazbayev, S. (2017). Sequential order as an extraneous factor in editorial decision. *Scientometrics*, *113*(3), 1573–1592. https://doi.org/10.1007/s11192-017-2531-7

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, *72*(6), 407–418. https://doi.org/10.1037/h0022602

Persson, E., Barrafrem, K., Meunier, A., & Tinghög, G. (2019). The effect of decision fatigue on surgeons' clinical decision making. *Health Economics*, *28*(10), 1194–1203. https://doi.org/10.1002/hec.3933

Philpot, L. M., Khokhar, B. A., Roellinger, D. L., Ramar, P., & Ebbert, J. O. (2018). Time of day is associated with opioid prescribing for low back pain in primary care. *Journal of General Internal Medicine*, *33*(11), 1828–1830. https://doi.org/10.1007/s11606-018-4521-8

Polman, E., & Vohs, K. D. (2016). Decision fatigue, choosing for others, and self-construal. *Social Psychological & Personality Science*, *7*(5), 471–478. https://doi.org/10.1177/1948550616639648

Ramji-Nogales, J., Schoenholtz, A. I., & Schrag, P. G. (2007). Refugee roulette: Disparities in asylum adjudication. *Stanford Law Review*, *60*(2), 295–411.

Read, D., & Loewenstein, G. (1995). Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology: Applied*, *1*(1), 34–49. https://doi.org/10.1037/1076-898X.1.1.34

Rotthoff, K. W. (2015). (Not finding a) sequential order bias in elite level gymnastics. *Southern Economic Journal*, *81*(3), 724–741. https://doi.org/10.1002/soej.12002

Salas, E., Rosen, M. A., & DiazGranados, D. (2010). Expertise-based intuition and decision making in organizations. *Journal of Management*, *36*(4), 941–973. https://doi.org/10.1177/0149206309350084

Scheer, J. K. (1973). Effect of placement in the order of competition on scores of Nebraska high school students. *Research Quarterly. American Association for Health, Physical Education and Recreation*, *44*(1), 79–85. https://doi.org/10.1080/10671188.1973.10615179

Shanteau, J. (1995). Expert judgment and financial decision making. In B. Green (Ed.), *Risky Business: Risk Behavior and Risk Management* (pp. 16–32). Department of Psychology, Kansas State University.

Shockley, K. M., Gabriel, A. S., Robertson, D., Rosen, C. C., Chawla, N., Ganster, M. L., & Ezerins, M. E. (2021). The fatiguing effects of camera use in virtual meetings: A within-person field experiment. *Journal of Applied Psychology*, *106*(8), 1137–1155. https://doi.org/10.1037/apl0000948

Sievertsen, H. H., Gino, F., & Piovesan, M. (2016). Cognitive fatigue influences students' performance on standardized tests. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(10), 2621–2624. https://doi.org/10.1073/pnas.1516947113

Simonsohn, U., & Gino, F. (2013). Daily horizons: Evidence of narrow bracketing in judgment from 10 years of M.B.A. admissions interviews. *Psychological Science*, *24*(2), 219–224. https://doi.org/10.1177/0956797612459762

Simonson, I. (1990). The effect of purchase quantity and timing on variety-seeking behavior. *Journal of Marketing Research*, *27*(2), 150–162. https://doi.org/10.1177/002224379002700203

Solomon, R. L. (1948). The influence of work on behavior. *Psychological Bulletin*, *45*(1), 1–40. https://doi.org/10.1037/h0055527

Sonnentag, S., & Zijlstra, F. R. H. (2006). Job characteristics and off-job activities as predictors of need for recovery, well-being, and fatigue. *Journal of Applied Psychology*, *91*(2), 330–350. https://doi.org/10.1037/0021-9010.91.2.330

Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, *54*(6), 377–411. https://doi.org/10.1037/h0043680

Stewart, N. (2009). Decision by sampling: The role of the decision environment in risky choice. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *62*(6), 1041–1062. https://doi.org/10.1080/17470210902747112

Sunstein, C. R., Kahneman, D., Schkade, D., & Ritov, I. (2001). Predictably incoherent judgments. *Stanford Law Review*, *54*(6), 1153–1215. https://doi.org/10.2307/1229618

Trougakos, J. P., & Hideg, I. (2009). Momentary work recovery: The role of within-day work breaks. In S. Sonnentag, P. L. Perrewé, & D. C. Ganster (Eds.), *Current perspectives on job-stress recovery* (pp. 37–84). Emerald Group Publishing. https://doi.org/10.1108/S1479-3555(2009)0000007005

Unkelbach, C., & Memmert, D. (2008). Game management, context effects, and calibration: The case of yellow cards in soccer. *Journal of Sport & Exercise Psychology*, *30*(1), 95–109. https://doi.org/10.1123/jsep.30.1.95

Unkelbach, C., & Memmert, D. (2014). Serial-position effects in evaluative judgments. *Current Directions in Psychological Science*, *23*(3), 195–200. https://doi.org/10.1177/0963721414533701

Unkelbach, C., Ostheimer, V., Fasold, F., & Memmert, D. (2012). A calibration explanation of serial position effects in evaluative judgments. *Organizational Behavior and Human Decision Processes*, *119*(1), 103–113. https://doi.org/10.1016/j.obhdp.2012.06.004

van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: Effects on perseveration and planning. *Acta Psychologica*, *113*(1), 45–65. https://doi.org/10.1016/S0001-6918(02)00150-6

Van Zant, A. B. (2022). Strategically overconfident (to a fault): How self-promotion motivates advisor confidence. *Journal of Applied Psychology*, *107*(1), 109–129. https://doi.org/10.1037/apl0000879

Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., & Tice, D. M. (2008). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology*, *94*(5), 883–898. https://doi.org/10.1037/0022-3514.94.5.883

Weinshall-Margel, K., & Shapard, J. (2011). Overlooked factors in the analysis of parole decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(42), Article E833. https://doi.org/10.1073/pnas.1110910108

Weiss, D. J., & Shanteau, J. (2003). Empirical assessment of expertise. *Human Factors*, *45*(1), 104–116. https://doi.org/10.1518/hfes.45.1.104.27233

Wickham, H. (2016). Data analysis. In H. Wickham (Ed.), *ggplot2: Elegant graphics for data analysis* (pp. 189–201). Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4_9

Wilson, V. E. (1977). Objectivity and effect of order of appearance in judging of synchronized swimming meets. *Perceptual and Motor Skills*, *44*(1), 295–298. https://doi.org/10.2466/pms.1977.44.1.295

Yariv, L. (2002). *I'll see it when I believe it? A simple model of cognitive consistency* (Cowles Foundation Discussion Paper No. 1352). https://ssrn.com/abstract=300696

Received August 18, 2020
Revision received September 30, 2022
Accepted October 2, 2022 ■