A Test of the Status Quo Theory of Depletion: Evidence from the Israeli Parole Decisions*

Øystein Daljord†, Oleg Urminsky‡& Jose-Manuel Ureta§

August 22, 2019

Abstract

Fatigue and depletion are decision making phenomenons that have received wide interest in consumer behaviour research. In a field study that is much cited in this literature, Danziger et al. (2011a) conjectured that the strongly cyclical parole decisions of an Israeli court could be explained by a status quo theory of depletion. As judges tire over the course of the day, it the judges would be more inclined to reach the decision of least resistance, which is to maintain the status quo by denying paroles. Danziger et al. (2011a) did not test its conjecture. We show that the empirical results in Danziger et al. (2011a) strongly refutes the status quo theory. The evidence is more mixed when we allow for judge specific depletion processes. We however find little support for the status quo theory of depletion overall. The declining parole frequencies may be consistent with a more general theory of depletion or fatigue that allows for a broader range of behaviour.

1 Introduction

The effects of fatigue have been widely studied in research on consumer decision-making. Consumer behavior researchers have specifically appealed to the resource-based ego depletion theory of consumer self-regulation (e.g., Baumeister (2002b)) to explain a wide range of findings, including effects of prior tasks on impulse buying (Vohs and Faber (2007)), susceptibility to context effects (Pocheptsova et al. (2009)), processing of advertising messages (Wheeler et al. (2007), Agrawal and Wan (2009), Wen Wan et al. (2010), reactions to product placement (Gillespie et al. (2012)) and intentions to purchase counterfeit products (Kim et al. (2012)). More generally, the theory of resource-based ego depletion has been incorporated into broader theories of consumer decision-making (Wang et al. (2010), Stilley et al. (2010a), Stilley et al. (2010b), Tour-Tillery and Fishbach (2012), Usta and Haubl (2011), Walsh (2014)).

Despite the increasing concerns about the soundness of evidence for the resource-based ego depletion theory, evidence has continued to accumulate that fatigue relates in some way to decision-making, with potentially dramatic consequences. Similar findings have been documented in consumer decision contexts, e.g. Levav et al. (2010). Recent research has likewise found that when patients have medical meetings later in the day, they are less likely to get breast and colorectal cancer screening (Hsiang et al. (2019)) and are more likely to be prescribed antibiotics (Linder et al. (2014)) and opioids (Philpot et al. (2018)). These empirical regularities are broadly consistent with the time-of-day trends in decisions that Danziger et al. (2011a) documented, and which are the focus of our study.

The studies of fatigue and depletion also tie into a larger debate in the social sciences

*We thank Jonathan Levav for the data and helpful discussions. Thanks to Simon Blanchard and Alexander Imas for comments. Financial support from the University of Chicago Booth School of Business is gratefully acknowledged.

†Booth School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA. E-mail: oeystein.daljord@chicagobooth.edu

‡Booth School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA. E-mail: oleg.urminsky@chicagobooth.edu

§Marine Farm, Chile.

about the degree to which humans succeed or fail to approximate optimal decisions. Much of the debate has centered on the question of whether people, particularly trained and experienced experts, are affected by extraneous factors that are normatively unrelated to optimal decisions, such as the sequencing of the decisions.[1]

Danziger et al. (2011a) studied expert decision making using sentencing data from an Israeli court of parole. It documented a striking pattern in the parole decisions. Its main result is replicated in Figure 1, which plots the relative parole frequency against the order the cases were heard over the course of a day. The relative parole frequency is strongly cyclical. Within each of three sessions, the initial parole frequency is about 60%, and declines to about zero by the end of the session. Following a snack break between sessions, the frequency rebounds and then gradually declines again towards the end of the session.

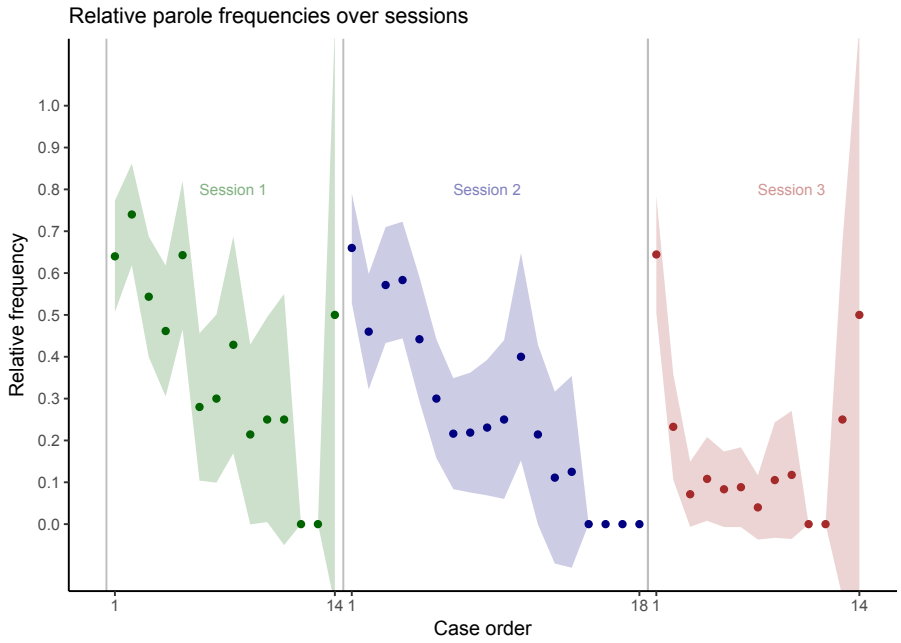

Figure 1: Relative parole frequencies of granted paroles plotted against case order, conditional on at least two observations per case order in each session. The relative parole frequencies are pooled across judges.

What makes the pattern informative is Danziger et al.'s claim that the case ordering was effectively randomized. Assuming that the case ordering is independent of the case characteristics, the null hypothesis of insensitivity to extraneous factors makes a sharp prediction: the parole frequency should be constant over the course of a session. It can be seen directly from Figure 1, and it is confirmed by the statistical analysis, that the parole frequencies are anything but constant.

Though Danziger et al. focused on rejecting the null hypothesis of insensitivity to extraneous factors, it conjectured that the declining frequencies within session and the spikes between sessions can be explained by a status quo theory of psychological depletion. While Danziger et al. did not explicitly test its conjecture, the paper has been interpreted as compelling field evidence for fatigue. in his NYT bestseller "Thinking fast and slow", page 44, Kahneman (2011) noted "The best possible account of the data provides bad news: tired and hungry judges tend to fall back on the easier default position of denying requests for parole. Both fatigue and hunger probably play a role.".

We identify new testable implications of the status quo theory. We give the statistical model in Danziger et al. an explicit sequential choice interpretation which allows us to test

[1] In the context of legal reasoning, legal formalists posit that judicial rulings are based exclusively on laws and facts, while judicial realists argue for a role of extra-legal psychological factors, see e.g. Leiter (2005) and Frank (1930).

the these implications. Our empirical approach is similar in spirit to the structural behavioral economics analysis of field experiments described in DellaVigna (2018) in that we state a set of assumptions that map the data distribution to parameters which characterize the depletion mechanism.

While the status quo theory does explain some of the patterns in the data, the theory also restricts the permissible range of some of these parameters. We find that under the assumptions of Danziger et al., the estimated parameters are inconsistent with the theory to a high degree of statistical precision. In subsequent analyses that allow for judge specific depletion processes, the status quo theory is strongly rejected for only a subset of the judges. Overall, we however find little empirical support for depletion.

When people are more fatigued, their cognitive performance declines, as documented in both lab-based accuracy tasks (Lorist et al. (2000)) and real-world tasks, such as driving performance (Campagne et al. (2004)). In particular, mental fatigue impedes people's success in goal-directed behavior, by shifting the balance between perceived reward and subjective cost of cognitively effortful activities, see e.g. Boksem and Tops (2008) for a detailed review. However, compared to the status-quo theory of resource depletion, broader theories of fatigue do not assume that the causes and consequences of fatigue are necessarily the same. We do not directly test our alternative account of fatigue, since it can support a comparatively larger range of *a priori* possible patterns of outcomes. We briefly discuss this point in Section 4.

1.1 Status Quo Depletion

A long history of research, going back to the 1980s, has documented systematic effects of fatigue on cognitive performance (Boksem and Tops (2008)). Recently, an influential literature in psychology has proposed that decision-making involves a limited mental resource, which is depleted by making choices, particularly choices requiring self-control (Baumeister et al. (2006), Vohs et al. (2008)). In this view, when the resource is depleted, people will be more likely to make choices that use less of the resource. In particular, this decision process was proposed to be both theoretically and empirically distinct from the effects of prior effort and fatigue (Baumeister et al. (2006)).

More specifically, the status quo depletion theory posits that not only is decision making itself cognitively taxing, but changing the status quo requires more justification, and hence more mental resources, than maintaining the status quo. In the context of Danziger et al.'s data, paroling a prisoner is considered changing the status quo, while denying a parole is considered maintaining the status quo. Therefore, as a judge is depleted over the course of a decision making session, the judge becomes more likely to deny paroles, either due to lack of mental resources or in a strategic attempt to conserve scarce remaining mental resources (Baumeister (2002a)).

According to the theory, depleted mental resources can be replenished by various factors including rest, food intake, and experiences that induce a positive mood (Tyler et al. (2008), Tice et al. (2007), Hagger et al. (2010)). If judges are replenished after a break, they are again prepared to exert mental effort and make decisions that change the status quo. Subsequently, the judge's mental resources are depleted by decision making over the course of the sessions and the parole propensity declines. Danziger et al. is careful to note that its data do not include measures directly related to depletion, such as blood sugar levels or a record of what the judges do in the breaks. Still, the paper argues that the breaks between sessions are plausibly used for activities which are shown to be replenishing in the depletion literature. We do not contest this assumption.

The notion of depletion as a meaningful decision making concept has lately been called into question (Inzlicht et al. (2014)). Meta-analytic reviews of lab-based tests of depletion have concluded that the effects are small and weak (Carter et al. (2015)). A recent high-powered pre-registered replication largely failed to replicate the lab-based findings (Hagger et al. (2016)), although there has been debate about the interpretation of this non-replication (Dang et al. (2016)). Given the current state of the literature, Danziger et al.'s evidence conjecture that status quo depletion explain the declining parole frequencies has been taken as unique

field-based evidence for the theory.

A feature of the status quo depletion theory which has gone unnoticed, is that the set of observable decisions that is consistent with the theory is restricted beyond declining parole propensities and spikes between sessions. The status quo depletion theory also implies that the current decision is influenced by both the number and the kind of decisions made so far in the session. In fact, the most common lab paradigms for testing depletion involve comparing performance in a second task among people who did a task predicted to be less depleting vs. those who did a task predicted to be more depleting.

As applied to the court context, a judge who has just granted a parole should be more depleted, and hence less likely to grant a subsequent parole than that judge would be if he had just denied a parole. According to the theory, the current propensity of parole therefore should depend negatively on the number of paroles granted so far in a session. This prediction is testable and the focus of our analysis in Section 2.2.

1.2 Analysis of judges' sentencing decisions

To test for the influence of extraneous factors, Danziger et al. presented both graphical evidence, in its Figure 1, and a series of logistic regressions (its main results are in its Table 1). In particular, the paper finds significant within-session correlation between the parole frequencies and the case order, as predicted by the status quo depletion theory.

Danziger et al.'s suggestion that depletion explains the decision patterns has been controversial for a variety of reasons. Under its assumption that the order of cases is drawn independently of the case characteristics, the effect of case order effects present dramatic evidence for legal realism, and against legal formalism. If the declining parole frequencies are indeed due to depletion, it is a real life example with grave social consequences of a controversial decision-making phenomenon that until then had been mostly demonstrated in the lab. Though these are important concerns from a legal perspective, our concern is not the consequences of the judges decisions for theories of justice, but testing the depletion theory itself.

The size of the estimated effects are large. The frequency of paroles drops about fifty percentage points over the course of a session. Glockner (2016) notes that the Cohen's d of 1.96 in Danziger et al. is about eight times higher than what had previously been found in the literature on depletion. The effect size raises legitimate concerns that the patterns may have other causes. One such concern is unobserved, case relevant factors that influenced the case ordering. Weinshall-Margel and Shapard (2011) argues that prisoners without legal representation are typically heard towards the end of a decision session. Weinshall-Margel and Shapard conducts interviews with judges that suggests there may be such systematic patterns in the ordering of cases. If so, the cyclical frequencies may emerge independently of depletion.

Danziger et al. (2011b) report data that do not show a systematic pattern in prisoners' legal representation. It also reports various institutional factors that may effectively randomize the case ordering. Though Danziger et al. (2011b) cannot rule out a systematic ordering of the cases, Section 2.2 shows that time varying factors are only weakly correlated with the observed case characteristics. Under some assumptions, which we discuss in Section B, correlations between the observable characteristics and the extraneous factors are informative about selection bias. We do not find such correlations in the data. Absent direct evidence to the contrary, the assumption of randomized case ordering is however untestable in our data.

Glockner (2016) offers an alternative account that does not appeal to decision biases. It shows that a particular set of assumptions on how judges decide when to take breaks can generate declining parole frequencies of similar magnitudes. Though we think there is merit to explore the explanatory power of rational models of decision making, our concern in this paper is not to provide an alternative way of rationalizing the data, but to test Danziger et al.'s conjecture. We however briefly discuss alternative accounts in Section 4.

We agree that the size of the effects in Danziger et al. (2011a) call into question whether depletion is the sole cause of the declining parole propensities. However, we do not discard

the status quo theory *a priori*. Instead, our approach is to restrict the analysis to only use the available evidence, taking the theory under consideration as given, and test more of the predictions of the proposed theory. All our analyses are premised on the case order being drawn independently of the case characteristics. Given that the case ordering is effectively random, we test whether the conjectured status quo depletion theory is consistent with the data.

2 An empirical model of status quo depletion

To the best of our knowledge, the status quo theory has only been described verbally in the literature. In this section, we show that the statistical models Danziger et al. uses to test for extraneous factors can be given a behavioural interpretation that allows us to test the status quo theory of depletion.

Danziger et al. uses logistic regressions to explain the parole decisions for each judge i, for each session s, and case order within session t, by a set of covariates. The covariates can be partitioned into a set of extraneous factors x and a set of case-relevant characteristics observable to the researchers z. The extraneous factors x used in Danziger et al. are the case order within the session (*caseorder*), the cumulative minutes spent in the current session (*cumulativeminutes*), and the proportion of paroles granted so far in the session (*pastparoles*). The case-relevant characteristics z include the nationality of the prisoner, the severity of the offense, and whether or not the prisoner had attended a rehabilitation program. The summary statistics of the data are given in the appendix.

The observed decisions are $y_{ist} \in \{0, 1\}$, representing parole denied and granted, respectively. The logistic regressions are of the kind

$$y^*_{ist} = \alpha_i + x_{ist}\beta_i + z_{ist}\gamma + \epsilon_{ist} \tag{1}$$
$$y_{ist} = \begin{cases} 1 & \text{if } y^*_{ist} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

where y^*_{ist} is latent index function. The error term ϵ_{ist} is a catch-all variable representing all factors that influence the decisions and that are unobservable to the researcher. The error term is assumed to be *iid* extreme value distributed. The null hypothesis of no influence of extraneous factors is a test of the restriction that β_i, the coefficients on the extraneous factors, are all zero. Danziger et al. (2011a)'s report results under the restriction that these parameters are equal across judges, i.e. $\beta_i = \beta$. Figure 1 shows that the null is rejected beyond conventional levels.

The statistical model in (1) can be given an explicit behavioural interpretation. We can think of the model as representing the decision process of judges who take the sum of all factors into account, where the parameters β and γ are weights which the judges attach to the observable extraneous and relevant factors, respectively. If the weighted sum of the four factors exceeds a threshold, without loss of generality set equal to zero, the judge grants the parole. Otherwise, the judge denies the parole.

The threshold can equivalently be thought of as varying with the extraneous factors x_{ist}, such that a parole is granted if $\alpha_i + z_{ist}\gamma + \epsilon_{ist} > -x_{ist}\beta_i$. Since the extraneous factors include covariates that vary over the session, the model allows the decision threshold to vary over the course of a session, consistent with a depletion process.

We next make a set of substantive assumptions that allows us to interpret the parameters of the statistical models as characterizing a status quo depletion process.

Assumption 1. *Each judge has a time invariant leniency threshold.*

We interpret the judge specific, time invariant fixed effect α_i in (1) as a judge specific disposition for leniency. Between sessions, the judges' mental resources are assumed to be replenished to the judge specific average leniency level. We distinguish between the leniency, which can be thought of as the mental resource capacity, and the depletion process, which can be thought of as mental resource taxation.

Assumption 2. *The depletion process depends on the number of decisions made so far in the session, the kind of decisions made so far in the session, and the time spent in the session.*

We represent the depletion process by the linear index $[pastparoles_{ist}, caseorder_{ist}, cumulativeminutes_{ist}]\beta_i$. This is similar to the specification in Danziger et al., but we substitute the number of past paroles in session, $pastparoles_{ist} = \sum_{\tau=0}^{t-1} y_{is\tau}$, in for the proportion of past paroles in session used in Danziger et al. as the function of past decisions. Our parameterization reflects the assumption that depletion is a cumulative, rather than relative phenomenon. Changing the variable representing past choices has little impact on the results of our tests, but gives a cleaner interpretation. We discuss this choice further in Section 2.2.

The specifications we use impose substantive restrictions on the depletion process. Only the number of paroles granted so far in the session matter for the current decision, and not the ordering or the time they were granted. This exclusion restriction implies that the mental depletion following a decision is summarized by the decision itself.[2] While more general depletion processes can be imagined, we adopt a specification as close as possible to Danziger et al.'s that can also parsimoniously accommodate a differential impact of paroles and denials on the mental resources.[3]

Assumption 3. *The mental replenishment between sessions is independent of past depletion.*

Assumption 4. *The depletion process is invariant across sessions.*

Assumption 3 implies that there is no systematic carry-over of depletion from decisions in previous sessions and sets the initial condition $y_{is0} = 0$, for all judges i and sessions s. The assumption breaks dependence between sessions, and by extension between days. Assumption 4 allows us to pool across sessions, which saves on degrees of freedom.

Assumptions 3 and 4 seem largely consistent with the data. From Figure 1, the empirical parole frequency starts at about the same level and declines at about the same rate for the first two sessions. The last session however seems to start at a lower level and drop off at a steeper rate. The approximation bias that would result if the last session follows a different depletion process and replenishment level may however be balanced by a decrease in variance as we can pool observations across all sessions.

Whereas Danziger et al. uses a more flexible function of *caseorder* and *session* in their main specification to approximate the parole propensity path over the course of the sessions , we opt for a more parsimonious specification to accommodate heterogeneity across judges given our limited degrees of freedom.

Assumption 5. *The case ordering is drawn independently of case characteristics*

Assumption 6. *The unobservable factors ϵ_{ist} are drawn from a continuous distribution, independently of the extraneous factors x_{ist}*

The assumption of a continuous distribution of ϵ in Assumption 6, along with monotonicity implied by additively separability in 1, ensures that the index function can be uniquely recovered from the data up to a scale. The assumption is satisfied by the extreme value error distribution, which we use in the application.

Note that we do not require ϵ to be conditionally independent of the observable case characteristics z. If the case ordering is randomized, then z are also drawn independently of the extraneous factors x. Consistency of α and β, our parameters of interest, is therefore robust to dependence of ϵ given z, but consistency of the nuisance parameters γ is not.

Assumption 7. *The judges are myopic.*

[2] Conditioning on the history of case characteristics would require substantially more information to estimate the process to a reasonable level of precision.

[3] An alternative specification allows recent paroles to have a stronger impact on the current decision than past paroles. We have however found little guidance from neither theory or empirics which can inform this choice. The results are qualitatively fairly robust to such alternative specifications and our chosen specification has the virtue of being more parsimonious.

Assumption 7 prohibits judges from strategically saving resources in expectation of future decisions. It assumes away potentially testable implications of some depletion theories. For instance, Baumeister (2002a) argues that part of the reduction in self-control is strategic, such that decision-makers refrain from exercising self-control in order to conserve the limited resource for future decisions.

The testable implication of this claim is that people will be less likely to choose the status quo option when they are making their last decision before an opportunity to replenish (e.g., before a break or at the end of the day), as long as they know it is the last decision. This claim has some empirical support. Muraven (1998) reports finding more of a depletion effect (i.e. reduction in self-control) on a second task when participants know a third task is coming, compared with when they are not aware of the third task.

Accommodating strategic self-control requires substantial additional assumptions on how judges form expectations over the future cases and how judges decide to when to end a session. Though there seems to be an uptick in the parole propensity towards the end of first and the third session in Figure 1, these upticks are based on only a handful of observations. Given the limited data and the need for a parsimonious model that can be tested with the available data, we defer consideration of this prediction to future research. Moreover, we do not explicitly account for how judges decide to end a session, which reportedly is at their discretion, but take the number of cases heard in each session as given. Accounting for the choice of when to take a break, while arguably important (Glockner (2016)), substantially complicates the analysis and is deferred to future research.

We estimate two models, corresponding to either of the following two assumptions.

Assumption 8. *The depletion processes are homogenous across judges.*

Assumption 9. *The depletion processes are judge specific.*

Assumption 8 restricts each judge to find it equally depleting to either grant or deny a parole, which implies the parameter restrictions $\beta_i = \beta$. We call this the homogenous depletion process assumption. Assumption 9 leaves all depletion process parameters β_i judge specific. We call this the heterogenous process assumption.

2.1 Inference

The empirical model in (1) is an unbalanced panel, binary choice model with lagged dependent variables and fixed effects. Given Assumptions 3 and 4, which ensures independence between sessions, we take the number of judges to be fixed and consider sessions the relevant asymptotic dimension, i.e. a fixed N large T panel. The lagged dependent variables are not strictly exogenous, which creates error correlations within sessions. We therefore cluster the standard errors on the session level.

2.2 Testing the status quo theory

The status quo theory makes three main predictions. The first is that the parole propensities are declining within session. The second is the parole propensities spike after breaks. The third is that past paroles has a negative influence on the current parole propensity.[4] Collectively, these three predictions allow a larger set of outcomes than the null of no extraneous factors, but they still rule out a substantial set of plausible outcomes as inconsistent with the theory. It is sufficient for one of the three predictions to fail to reject the theory. From Figure 1, the first two predictions are clearly consistent when the data are pooled across judges.

It is generally harder to detect the third prediction, negative (conditional) state dependence, graphically. Danziger et al.'s regression analyses however include a test of state dependence. The majority of the statistical models reported in Danziger et al. include a variable that records the proportion of past paroles that day. This variable was intended to control for a possible quota of paroles, a decision making phenomenon in which the judge sets

[4]These three predictions may not exhaust all restrictions the status quo theory imposes on the data. Without a precisely stated theory, it is however difficult to determine the full range of testable implications.

a target number of paroles to grant within a session.[5] However, as a function of past decisions, the variable captures state dependence in the decisions and belongs to the set of testable extraneous factors.

Decisions are state dependent when the current decision is influenced by decisions in the past. Positive state dependence results in runs of similar decisions. One well-known example of positive state dependence is the 'hot hand' in basketball. Negative state dependence is the opposite. A similarly well known of (illusionary) negative state dependence is the 'gamblers fallacy', where people tend to perceive that a heads of a fair coin is more likely to be followed by tails than heads. Across all 15 reported specifications in Danziger et al., the parameter on the proportion of past paroles is positive and mostly highly statistically significant. Our homogenous depletion process specification, which is reported in Section 3, replicates this finding.

Positive state dependence implies that the more paroles a judge has granted so far in that session, the more likely he is to grant another parole, *all else equal*.[6] The qualifier is important. As seen in Figure 1, the parole frequency is clearly declining within session in Figure 1. By "all else equal", we mean that these are runs of similar decisions relative to that declining trend. Hence, positive state dependence is consistent with declining parole propensities, but inconsistent with the status quo theory, which only allows negative state dependence. Therefore, Danziger et al.'s finding of positive state dependence rejects the status quo theory.[7]

We next note that although the depletion theory as stated in Baumeister and Muraven (2000) suggests heterogeneity in depletion processes, the statistical models in Danziger et al. imply homogenous depletion processes. Allowing for heterogenous depletion processes raises three separate issues. Firstly, we must now consider the domain of the theory: does it apply to all judges, or just some? If it applies to more than one judge, which ones? On this issue, we find limited guidance from theory. Following Glockner and Betsch (2011), we choose to maximize the empirical content by interpreting the theory as predicting that all judges display a pattern consistent with the theory. This is stronger than typical tests of behavioural theories in the lab where average effects are typically the statistics of interest. We however think that a conjecture can be as interesting in what it fails to explain as in what it does explain. In the end, we think of rejecting a falsifiable theory as progress, rather than failure.[8]

Secondly, the heterogenous specification comes at the cost of degrees of freedom which results in a substantial loss of power relative to the homogenous model. We show in Section 3 that the heterogenous specification yields mixed for the status quo theory, but the evidence is also noisy. Though we do test and reject the theory that all judges are subject to status quo depletion, we view these results as more exploratory and display the results for all judges individually. Thirdly, restricting the specification to homogenous depletion processes when these processes are heterogenous, leads to measurements issues for our key test statistic (Heckman (1981)), which we discuss in the following section.

2.3 Measuring state dependence

State dependence is a common feature in panels of choice data in marketing and economics, ranging from demand for consumer goods, e.g. Seetharaman et al. (1999), to employment choices, e.g. Keane and Sauer (2009). An important question in this literature is whether the state dependence is *structural* or *spurious*, see e.g. Heckman (1981). Structural state dependence reflects a decision making phenomenon, while spurious state dependence is caused

[5] It is unclear how the proportion of past decisions can capture a target number of paroles.

[6] Danziger et al. notes this fact on pp. 6892: "Regardless of the analysis we conducted, the parameter estimate was positive and significant, suggesting that a judge who made a large proportion of favorable rulings up to a certain point was, in fact, more likely to rule favorably in a subsequent case.", but does not discuss it further.

[7] Chen et al. (2016) finds strong and significant negative state dependence in the sequential decisions of baseball umpires, loan application officers, and asylum court judges. It takes the evidence as consistent with the 'gamblers fallacy', but seemingly does not find evidence of spikes and declines between the sessions.

[8] The field of judgment and decision making research has indeed emerged from repeatedly rejecting falsifiable theories of various characterizations of rational choice.

by a failure to account for unobserved heterogeneity. In our case, unobserved heterogeneity means that the depletion processes are judge specific. Whether we can interpret the estimated state dependence as informative about the depletion process therefore depends on whether we can interpret it as structural.

A structural interpretation of the state dependence parameter in the regressions in Danziger et al. requires assuming that all judges follow the same depletion process, except for judge specific leniencies. Since the parameters on all the extraneous factors are homogenous across judges in Danziger et al.'s main specification, unobserved heterogeneity in the depletion process leads to residuals that are positively correlated over time within judge. This positive correlation in the residuals loads onto the variable of past decisions and shows up as positive state dependence in the model. The estimated state dependence may therefore reflect a failure of the model to account for unobserved heterogeneity in the decision process, which is spurious, rather than then reflect that paroles are more depleting than denials, which is structural. We give an illustration of the phenomenon in Appendix A.

It is generally difficult to distinguish between structural and spurious state dependence in the data. We adopt the approach of Heckman (1981) and Dubé et al. (2010). The idea is simple: while structural state dependence survives in specifications with richer heterogeneity, spurious state dependence may not. We therefore interpret the state dependence that remains after controlling for heterogeneity as structural. We report one homogenous specification and one heterogenous specification. Both satisfy Assumptions 1 to 7. The homogenous model also satisfies Assumption 8. It is similar to the analyses in Danziger et al. and replicates its findings. The heterogenous model satisfies Assumption 9.

3 Empirical results

The estimates of the two models are given in Table 1. For the homogenous model, the estimated state dependence parameter for $nrfavprsesssion$ is 0.36, notably positive, and for $caseorder$ is -0.42, both highly significantly different from zero. The results therefore reject the null of no influence of extraneous factors, which replicates the Danziger et al. result. The positive state dependence is also inconsistent with the homogenous version of the status quo depletion theory, as it implies that granting paroles is less depleting than denying paroles. Though there is independent variation in *minutes spent in session*, the marginal effect time of time spent in session is effectively zero, with small standard errors. To conserve degrees of freedom, we do not estimate judge specific effects on *minutes spent in session* in the heterogenous specification.

In Figure 2 we have plotted error bars for the estimated judge-specific depletion rates and the state dependence parameters from the heterogeneous model. The error bars are overlaid with a 95% confidence interval for the parameters from the homogenous specification (the grey box). Note that the estimated state dependence for judge i is the sum of coefficients on $pastparoles$ and $pastparoles_i$.

The judge specific depletion rates are all centered around the value of the pooled parameter from the baseline specification, at -0.42 (the purple line), but with substantial heterogeneity across judges. The fact that the state dependence parameters from the heterogenous specification are centered around 0.17, down from the precisely estimated value of 0.36 in the homogenous specification, suggests that part of the positive state dependence in the homogenous specification can be explained by unobserved heterogeneity.

For three of the eight judges (judges 4, 5 and 8), the estimated state dependence parameters are directionally negative, consistent with the status quo theory. The estimates are however imprecise, particularly for judge 4, and for judge 8, the parameter is virtually zero. The estimated state dependence parameters are directionally positive for the remaining five judges, which is inconsistent with the theory. Only judges 1 and 7 are significantly different from zero, with nominal p-values given as 0.00005 and 0.00048, respectively.

A test of the null hypothesis of the status quo depletion theory that restricts all judges to have non-positive state dependence, i.e. the strong theory, is rejected if at least one judge

	Homogeneous depletion process	Heterogenous depletion process
intercept	−0.54 (0.91)	−0.47 (0.99)
Extraneous factors		
pastparoles	0.36 (0.07)***	0.51 (0.13)***
caseorder	−0.42 (0.07)***	−0.53 (0.13)***
cumulativeminutes	0.00 (0.01)	0.01 (0.01)
Case characteristics		
offence	0.05 (0.10)	0.05 (0.10)
imprison	−0.25 (0.06)***	−0.24 (0.06)***
monthsrv	−0.00 (0.00)	−0.00 (0.00)
rehab	1.96 (0.87)*	2.12 (0.85)*
nationality	−0.14 (0.15)	−0.10 (0.15)
gender	−0.30 (0.33)	−0.41 (0.35)
Judge specific leniency		
judge 2	−0.06 (0.24)	−0.11 (0.58)
judge 3	−0.34 (0.28)	0.04 (0.50)
judge 4	−0.28 (0.24)	−0.05 (0.49)
judge 5	−0.12 (0.28)	0.92 (0.70)
judge 6	0.33 (0.29)	−0.03 (0.45)
judge 7	0.03 (0.29)	0.08 (0.54)
judge 8	0.20 (0.28)	−1.12 (0.48)*
Judge specific state dependence		
pastparoles judge 2		−0.20 (0.23)
pastparoles judge 3		−0.23 (0.25)
pastparoles judge 4		−0.66 (0.37)
pastparoles judge 5		−0.78 (0.62)
pastparoles judge 6		−0.36 (0.37)
pastparoles judge 7		0.05 (0.21)
pastparoles judge 8		−0.53 (0.27)*
Judge specific depletion rates		
caseorder judge 2		0.11 (0.21)
caseorder judge 3		−0.02 (0.17)
caseorder judge 4		0.15 (0.20)
caseorder judge 5		0.02 (0.32)
caseorder judge 6		0.25 (0.19)
caseorder judge 7		−0.03 (0.20)
caseorder judge 8		0.47 (0.17)**
Num. obs.	1112	1112

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 1: Statistical models

has a statistically significant positive state dependence parameter. This test consists of eight individual tests, one for each judge, so we need to account for multiple comparisons.[9] The family wise error rate is the probability of rejecting at least one true null hypothesis when conducting multiple tests. With a Bonferroni correction, which is a conservative bound for the family wise error rate, the 1%-critical nominal p-value is $\frac{0.01}{8} = 0.0013$, which is larger than the test statistics for judges 1 and 7 by a good margin.

It is still not entirely clear which family of tests we consider. If we consider a test of the strong theory as making three predictions for each judge (spikes, declines, and negative state dependence), we get a total of 24 tests. For this case, the Bonferroni correction gives a 1% critical value at nominal p-value $\frac{0.01}{24} = 0.0004$. Hence, the strong test rejects also this larger family of tests at the 1% level, and still by a good margin, irrespective of whether the prediction of spikes and declines are met in the data.

Instead of formally testing the predicted declining frequencies within sessions and spikes between sessions, we choose a more exploratory analysis of the patterns. While the results from the homogeneous specification show strong evidence of the predicted spikes, the results from the heterogenous specification are more mixed across judges. In Figure 5, the predicted parole propensities from the heterogenous specification are plotted for each judge. Judges 1 to 5 and Judge 7 display declining empirical frequencies within session and spikes between sessions,

[9] While the status quo theory restricts the state dependence parameter to be negative, our null hypothesis is that the state dependence parameter is exactly zero, the largest value that is consistent with the theory. In that sense, our test is conservative.

while Judges 6 and 8 do not display clear evidence of either declines or spikes.

One possibility is that the status quo depletion is a phenomenon that applies to some judges, but not others. The observed choices may be consistent with depletion at the aggregate level, even if not all judges exhibit status quo depletion individually. For example, extending the "self-control as a muscle metaphor", Baumeister and Muraven (2000) propose that people with more practice or experience may exhibit weaker depletion effects. This predicts that some judges would show declines during sessions, spikes between sessions, and negative state dependence, while other judges would show weak or no evidence for all three patterns.

However, the results from the judge-level analysis provide little support for this possibility. Only two of the judges (4 and 5) who show clear spikes in Figure 5 also have an estimated negative state dependence. The state dependence parameter is particularly imprecisely estimated for judge 5, which may reflect that we have low power. Judges 1 and 7 show clear spikes and declines, but these are the two judges with more precisely estimated positive state dependence, which point to a decision process incompatible with status quo depletion for these judges.

Judge 6 displays only modest declines and spikes. Judge 8 has a pattern that is clearly consistent with the null hypothesis of no extraneous factors: the depletion rate is close to zero, there are no obvious spikes between sessions, and there is no state dependence. Judge 4 may be said to display a pattern that meets all three predictions. Overall, the results therefore do not clearly align with the predictions of the status quo theory.

We emphasize again that as we are running out of degrees of freedom, we are limited in the positive conclusions that can be drawn. Throughout this analysis, we maintain the assumption of randomized case ordering. We cannot test this assumption directly, but we have not found systematic patterns in the observable case characteristics. We discuss some diagnostic tests that are common in the treatment effect literature in the appendix. These tests show no obvious signs of non-randomness. Still, absence of evidence is not evidence of absence. It is possible that there is some unobserved factor that influences both the order which the case is heard and the judges' decisions and which is uncorrelated with the observable characteristics. Finally, though we have accounted for multiple dimensions of heterogeneity, we can not rule out that there is a spurious component to the estimated state dependence that reflects residual unobserved heterogeneity.

4 Alternative accounts

A virtue of the status quo depletion theory of Danziger et al. is that it has specificity: it rules out a material subset of the possible outcomes as inconsistent with the theory. It is due to the specificity of the status quo depletion theory that we could test this controversial theory, rather than merely question the assumptions of the theory.

The specificity comes from the status quo feature of the theory – the assumption that the decision to grant parole, as a violation of the status quo, both impacts and is impacted by the degree of depletion. It restricts the parole propensities to decline within sessions and it rules out positive state dependence. Without the status quo feature, the only restrictions of more general depletion or fatigue theories on the parole decisions is a discontinuity in the parole propensity between sessions, and non-constant propensities within sessions.

These restrictions allow for a substantially wider range of deviations from a constant parole frequency, the null tested by Danziger et al., to be consistent with some depletion-like process. It is therefore harder to reject. Similarly, if we changed the meaning of status quo from keeping a prisoner incarcerated to something else, such as equating status quo with making the same choice as in the previous decision, the theory may again be consistent with the data. Adapting the theory to accommodate a wider range of data however empties the theory of empirical content.

It may very well be, as suggested by Kahneman, that hunger and fatigue can explain the patterns independently of, or perhaps in combination with, depletion, without restricting

either the sign of the state dependence or the time trend in the parole propensities. It is for example *a priori* plausible that simply deciding cases, whether or not parole is granted, is tiring. Fatigue has been shown to reduce attention and increase distractions and errors (Boksem and Tops (2008)). If fatigue in turn leads to less deliberation and greater weight on initial impressions (Webster et al. (1996)), fatigue may reduce the parole propensity, which is consistent with the results. However, fatigue could by the same arguments support increasing parole propensities. Meaningful tests of such theories may require additional data on the judges' mental states.

The example shows that finding an alternative theory that rationalizes known data is usually not that hard. As Glockner and Betsch (2011) notes, rationalizing a particular outcome is not the only criterion of a good theory, it should also exclude a large range of outcomes as inconsistent with the theory. One would hope that alternative accounts have the specificity of the status quo theory. Moreover, a persuasive alternative theory would not only explain the known data, but make refutable predictions for a larger class of decision problems than the one at hand, and, in the best of all worlds, clearly specify which class of decision problems that is.

5 Summary

We developed a test of Danziger et al.'s suggested theory of the cyclical patterns in the Israeli judges parole decisions. We found that the status quo theory of depletion is strongly rejected by the positive state dependence in the observed choices. As unobserved heterogeneity in the depletion processes is known to cause spurious state dependence, we next allowed for judge specific depletion processes. Accounting for multiple dimensions of depletion process heterogeneity, we do not find strong support for the theory, either when assuming that all judges or a subset of the judges are subject to status quo depletion. Only one judge out of eight may be said to display a pattern that is consistent with the status quo theory, whereas two judges display patterns that are strongly inconsistent with the theory. For the remaining judges, we find the evidence is inconclusive when taking all predictions of the theory into account. We emphasize that our analysis was made possible by Danziger et al. suggesting a refutable explanation of the observed choice patterns. More general theories of depletion and fatigue that allow for a broader set of behaviours may be consistent with the data.

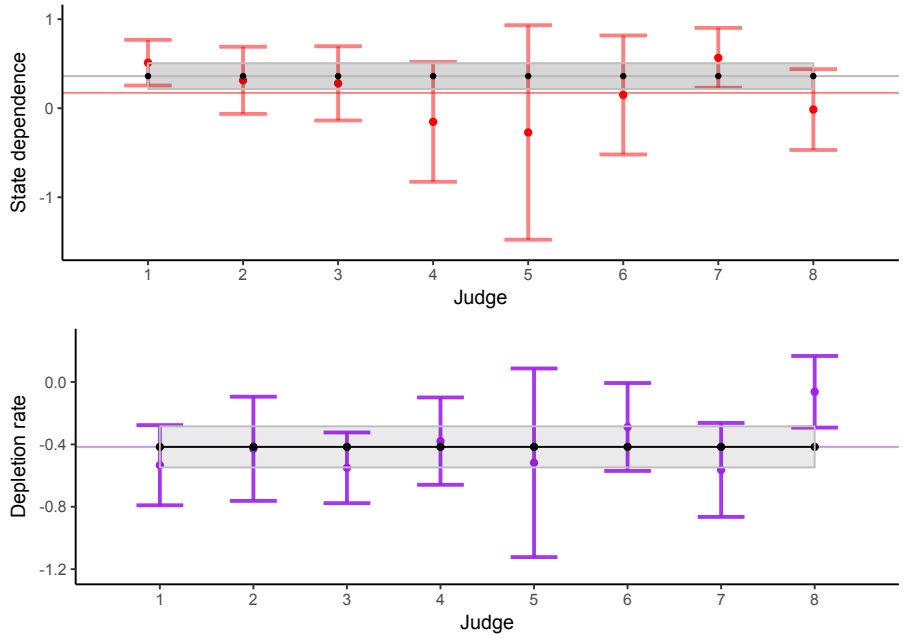

Figure 2: The grey shaded areas in the upper and lower plots give the 95% confidence interval from the pooled regression. The grey horizontal lines are set at the pooled point estimate. The coloured error bars give the 95% confidence intervals for the judge specific state dependence and depletion rates in the individual regressions. The coloured horizontal lines give the equally weighted average individual specific parameter.

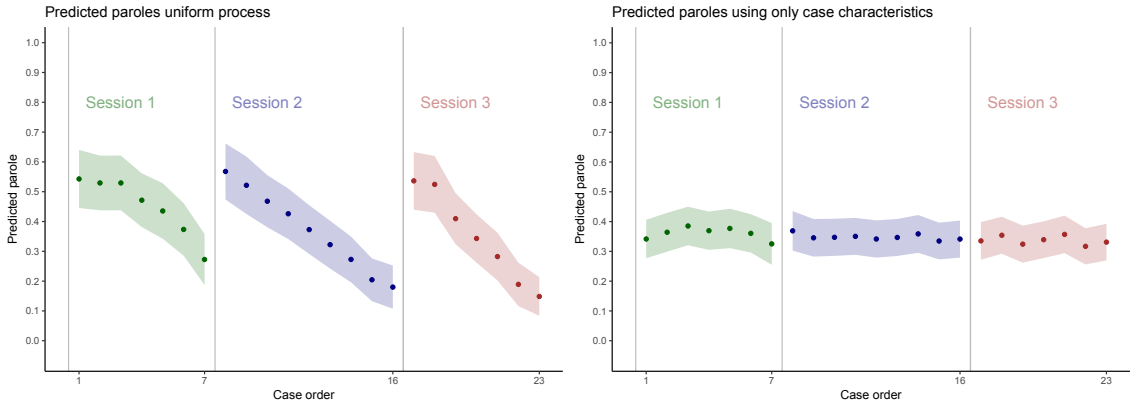

Figure 3: The predicted parole propensities are calculated at the average covariates for the average number of cases per session, assuming homogenous depletion processes, using the leniency (fixed effect) of judge 1.

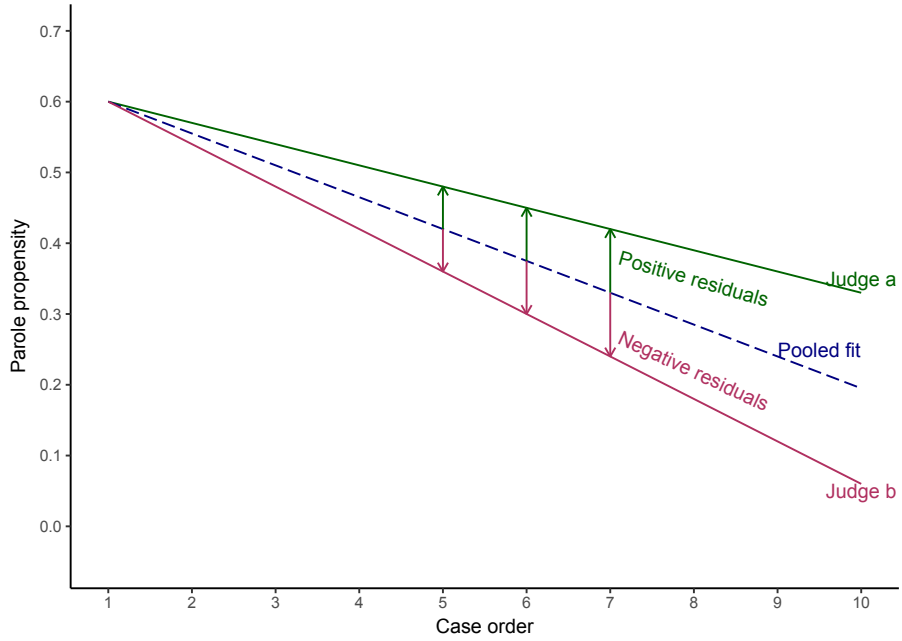

Figure 4: Example of positively correlated model errors resulting from heterogenous depletion processes across judges.

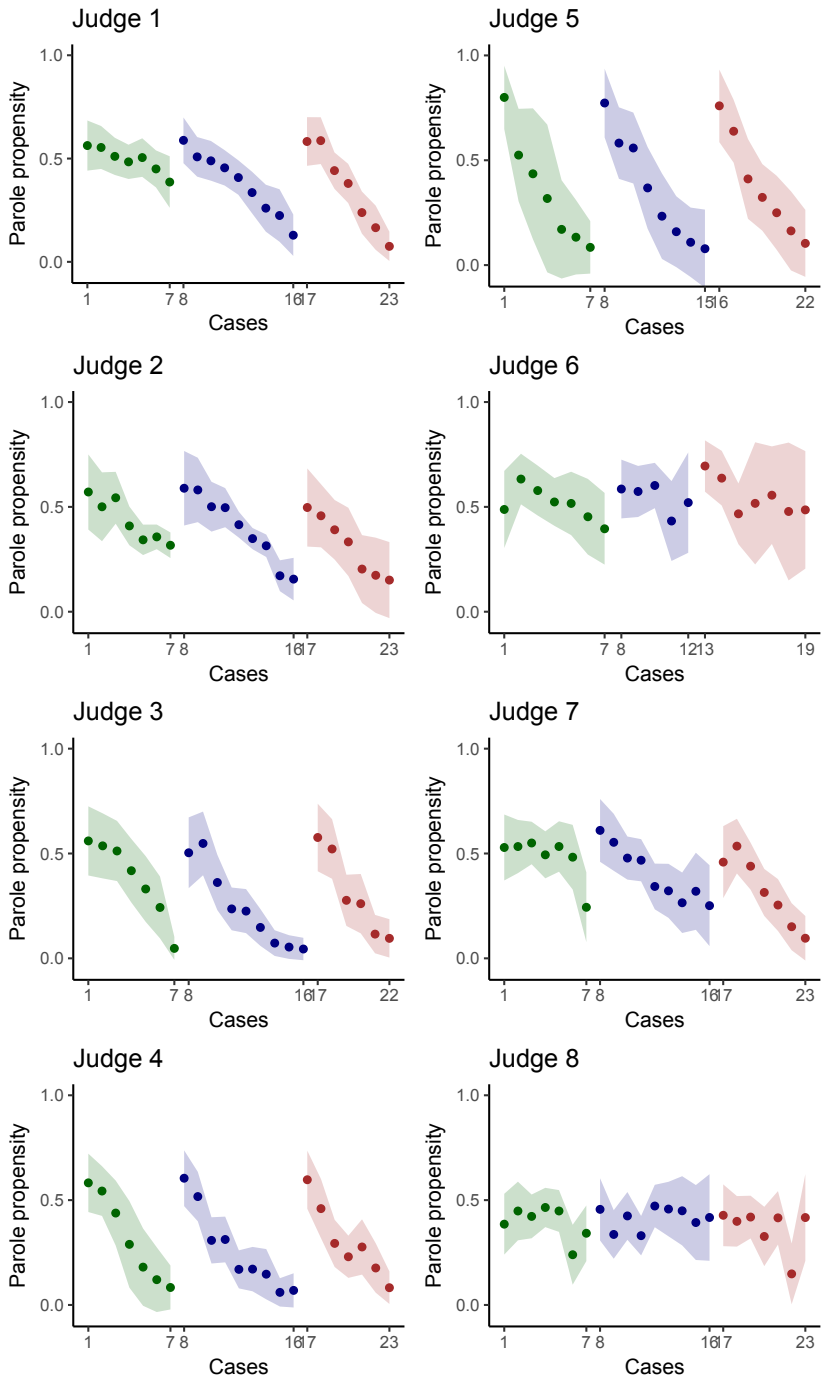

Figure 5: Predicted parole propensities with judge specific depletion rates and state dependence parameters. For each judge, the predicted propensities are calculated at the average covariates and for the average number of cases per session.

References

Agrawal, N. and E. W. Wan (2009, February). Regulating risk or risking regulation? construal levels and depletion effects in the processing of health messages. *Journal of Consumer Research 36*(3), 448–462. 1

Altonji, J., T. Elder, and C. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy 113*, 151–184. 1

Baumeister, R. (2002a). Ego depletion and self-control failure: An energy model of the self's executive function. *Self and identity 2*, 129–136. 3, 7

Baumeister, R., M. Gailliot, C. DeWall, and M. Oaten (2006). Self-regulation and personality: how interventions increase regulatory success, and how depletion moderates the effects of traits on behavior. *Journal of Personality 74*, 1773–1801. 3

Baumeister, R. and M. Muraven (2000). Self-regulation and depletion of limited resources: does self-control resemble a muscle? *Psychological Bulletin*, 247–59. 8, 11

Baumeister, R. F. (2002b, 03). Yielding to temptation: Self-control failure, impulsive purchasing, and consumer behavior. *Journal of Consumer Research 28*(4), 670–676. 1

Boksem, M. and M. Tops (2008). Mental fatigue: costs and benefits. *Brain Research Review 59*, 125–139. 3, 12

Campagne, A., T. Pebayle, and AlainMuzet (2004). Correlation between driving errors and vigilance level: influence of the driver's age. *Physiology & Behavior 80*(4), 515 – 524. 3

Carter, E., L. Koffler, D. Forster, and M. Cullough (2015). A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource. *Journal of Experimental Psychology 144*, 796–815. 3

Chen, D., T. Moskowitz, and K. Shue (2016). Decision making under the gamblers fallacy: evidence from asylum judges, loan officers, and baseball umpires. *Quarterly Journal of Economics 131*, 1181–1242. 8

Dang, J., S. Xiao, Y. Liu, and L. Mao (2016). Individual differences in dopamine level modulate the ego depletion effect. *International Journal of Psychophysiology 99*, 121–124. 3

Danziger, S., J. Levav, and L. Avnim-Pesso (2011a). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences USA 108*(17), 6889–6892. 1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12

Danziger, S., J. Levav, and L. Avnim-Pesso (2011b). Reply to Weinshall-Margel and Shapard: Extraneous factors in judicial decisions persist. *Proceedings of the National Academy of Sciences USA 108*(42), E834. 4

DellaVigna, S. (2018). Chapter 7 - structural behavioral economics. In B. D. Bernheim, S. DellaVigna, and D. Laibson (Eds.), *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 1 of *Handbook of Behavioral Economics: Applications and Foundations 1*, pp. 613 – 723. North-Holland. 3

Dubé, J.-P., G. Hitsch, and P. Rossi (2010). State dependence and alternative explanations for consumer inertia. *Rand Journal of Economics 41*(3), 417–445. 9

Frank, J. (1930). *Law and the Modern Mind*. New York: Brentano's. 2

Gillespie, B., J. Joireman, and D. D. Muehling (2012). The moderating effect of ego depletion on viewer brand recognition and brand attitudes following exposure to subtle versus blatant product placements in television programs. *Journal of Advertising 41*(2), 55–66. 1

Glockner, 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. 4, 7

Glockner, A. and T. Betsch (2011). The empirical content of theories in judgment and decision making. *Judgment and Decision Making 6*, 711–721. 8, 12

Hagger, M., N. Chatzisarantis, H. Alberts, C. Anggono, C. Batailler, A. Birt, R. Brand, M. Brandt, G. Brewer, S. Bruyneel, D. Calvillo, W. Campbell, P. Cannon, M. Carlucci, N. Carruth, T. Cheung, A. Crowell, D. De Ridder, S. Dewitte, M. Elson, J. Evans, B. Fay, B. Fennis, A. Finley, Z. Francis, E. Heise, H. Hoemann, M. Inzlicht, S. Koole, L. Koppel, F. Kroese, F. Lange, K. Lau, B. Lynch, C. Martijn, H. Merckelbach, N. Mills, A. Michirev, A. Miyake, A. E. Mosser, M. Muise, D. Muller, M. Muzi, D. Nalis, R. Nurwanti, H. Otgaar, M. Philipp, P. Primoceri, K. Rentzsch, L. Ringos, C. Schlinkert, B. J. Schmeichel,

S. Schoch, M. Schrama, A. Schtz, A. Stamos, G. Tinghog, J. Ullrich, M. vanDellen, S. Wimbarti, W. Wolff, C. Yusainy, O. Zerhouni, and M. Zwienenberg (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science 11*(4), 546–573. 3

Hagger, M., C. Wood, C. Stiff, and N. Chatzisarantis (2010). Ego depletion and the strength model of self-control: a meta-analysis. *Psychological Bulletin 136*, 495–525. 3

Heckman, J. J. (1981). Statistical models for discrete panel data. In C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 114–178. Cambridge, MA: MIT Press. 8, 9

Hsiang, E. Y., S. J. Mehta, D. S. Small, C. A. L. Rareshide, C. K. Snider, S. C. Day, and M. S. Patel (2019, 05). Association of Primary Care Clinic Appointment Time With Clinician Ordering and Patient Completion of Breast and Colorectal Cancer ScreeningClinic Appointment Time, Clinician Ordering, and Patient Completion of Cancer ScreeningClinic Appointment Time, Clinician Ordering, and Patient Completion of Cancer Screening. *JAMA Network Open 2*(5), e193403–e193403. 1

Inzlicht, M., B. Schmeichel, and C. Macrae (2014). Why self-control seems (but may not be) limited. *Trends in Cognitive Science 18*, 127–133. 3

Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux. 2, 11

Keane, M. and R. Sauer (2009). Classification error in dynamic discrete choice models: Implications for female labor supply behavior. *Econometrica 77*, 975991. 8

Kim, J., J.-E. Kim, and J. Park (2012). Effects of cognitive resource availability on consumer decisions involving counterfeit products: The role of perceived justification. *Marketing Letters 23*(3), 869–881. 1

Leiter, B. (2005). The blackwell guide to philosophy of legal theory. pp. 50–66. 2

Levav, J., M. Heitmann, A. Herrmann, and S. Iyengar (2010). Order in product customization decisions: Evidence from field experiments. *Journal of Political Economy 118*(2), 274–299. 1

Linder, J. A., J. N. Doctor, M. W. Friedberg, H. Reyes Nieva, C. Birks, D. Meeker, and C. R. Fox (2014, 12). Time of day and the decision to prescribe antibiotics. *JAMA Internal Medicine 174*(12), 2029–2031. 1

Lorist, M. M., M. Klein, S. Nieuwenhuis, R. de Jong, G. Mulder, and T. F. Meijman (2000). Mental fatigue and task control: planning and preparation. *Psychophysiology 37 5*, 614–25. 3

Muraven, M. (1998). Mechanisms of self-control failure: Motivation and limited resources. Unpublished doctoral dissertation. 7

Oster, E. (2016). Unobservable selection and coefficient stability. theory and evidence. Working paper, Brown University. 1

Philpot, L. M., P. Ramar, M. Y. Elrashidi, T. A. Sinclair, and J. O. Ebbert (2018). A before and after analysis of health care utilization by patients enrolled in opioid controlled substance agreements for chronic noncancer pain. *Mayo Clinic Proceedings 93*(10), 1431 – 1439. 1

Pocheptsova, A., O. Amir, R. Dhar, and R. F. Baumeister (2009). Deciding without resources: Resource depletion and choice in context. *Journal of Marketing Research 46*(3), 344–355. 1

Seetharaman, P. B., A. Ainslie, and P. Chintagunta (1999). Investigating household state dependence effects across cateogries. *Journal of Marketing Research 36*(4), 488–500. 8

Stilley, K. M., J. J. Inman, and K. L. Wakefield (2010a, 02). Planning to make unplanned purchases? the role of in-store slack in budget deviation. *Journal of Consumer Research 37*(2), 264–278. 1

Stilley, K. M., J. J. Inman, and K. L. Wakefield (2010b). Spending on the fly: Mental budgets, promotions, and spending behavior. *Journal of Marketing 74*(3), 34–47. 1

Tice, D., R. Baumeister, D. Shmueli, and M. Muraven (2007). Restoring the self: Positive affect helps improve self-regulation following ego depletion. *Journal of Experimental Social Psychology 43*, 379–384. 3

Tour-Tillery, M. and A. Fishbach (2012). The end justifies the means, but only in the middle. *Journal of Experimental Psychology 141*(3), 570–583. 1

Tyler, J. M., , and K. C. Burns (2008). After depletion: The replenishment of the self's regulatory resources. *Self and Identity 7*, 305–321. 3

Usta, M. and G. Haubl (2011). Self-regulatory strength and consumers relinquishment of decision control: When less effortful decisions are more resource depleting. *Journal of Marketing Research 48*(2), 403–412. 1

Vohs, K., R. Baumeister, B. Schmeichel, J. Twenge, N. Nelson, and D. Tice (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*, 883–898. 3

Vohs, K. D. and R. J. Faber (2007, 01). Spent resources: Self-regulatory resource availability affects impulse buying. *Journal of Consumer Research 33*(4), 537–547. 1

Walsh, D. (2014). Can priming a healthy eating goal cause depleted consumers to prefer healthier snacks? *Journal of Consumer Marketing 31*(2), 126–132. 1

Wang, J., N. Novemsky, R. Dhar, and R. F. Baumeister (2010). Trade-offs and depletion in choice. *Journal of Marketing Research 47*(5), 910–919. 1

Webster, D. M., L. Richter, and A. W. Kruglanski (1996). On leaping to conclusions when feeling tired: Mental fatigue effects on impressional primacy. *Journal of Experimental Social Psychology 32*, 181–195. 12

Weinshall-Margel, K. and J. Shapard (2011). Overlooked factors in the analysis of parole decisions. *Proceedings of the National Academy of Sciences USA 108*(42), E833. 4

Wen Wan, E., D. Rucker, Z. L Tormala, and J. J Clarkson (2010, 06). The effect of regulatory depletion on attitude certainty. *Journal of Marketing Research 47*, 531–541. 1

Wheeler, C., P. Brinol, and A. Herrmann (2007). Resistance to persuasion as self-regulation: Ego-depletion and its effects on attitude change processes. *Journal of Experimental Social Psychology 43*(1), 150 – 156. 1

A Example of spurious state dependence

Suppose two judges a and b both have parole propensities that decline over a decision session, as illustrated in Figure 4. The propensities decline at different rates for the two judges, but by assumption, there is no state dependence for either judge. Suppose next that we pool the observations for both judges and estimate a slope that is homogenous across judges. The fit from the pooled regression will be an average of the two slopes.

The residuals from the fitted specification will be autocorrelated, though the model errors are not. The residuals in the example will tend to be positive for judge a and negative for judge b, that is, the residuals are positively correlated. The model will underestimate the propensity for judge a (positive residuals) and overestimate the propensity for judge b (negative residuals). In both cases, the variation tends to be absorbed by included functions of past decisions, which gives spurious state dependence.

Similarly, we can note that judge a will grant more paroles than judge b over the course of a session. Judge a therefore has a higher stock of past paroles relative to judge b. This higher stock of past paroles is predictive of his future parole propensity, which is higher than the average. Similarly, judge b will grant fewer paroles, and his lower stock of past paroles will explain the negative errors. In both cases, the correlated errors load onto the variable of past decision. In this case, the recovered positive state dependence reflects unobserved heterogeneity in the depletion process, and not a distinct underlying decision making phenomenon.

B Correlations between the extraneous factors and the observable case characteristics

We do not formally test the assumption of randomized case ordering. Instead, we follow a common informal approach which tests if the structural parameters are sensitive to the inclusion of outcome relevant control variables. If the structural parameters are robust to inclusion of relevant control variables, then they are likely also robust to biases from omission of unobserved controls, such as variables that determine the ordering of cases. Though formal tests based on the approach exist, see Altonji et al. (2005) and Oster (2016), we do the informal one to keep it simple.

In Table 2, we have reproduced the homogenous model with case characteristics from Table 1 in the first column for convenience. In the second column, we have estimated the homogenous model without including the case characteristics. The state dependence and depletion rate parameters hardly move between the two models, though more than one of the observable characteristics are significant. Since adding the observable case relevant factors has little to no impact on the structural parameters of interest, but does significantly improve the fit of the model, one may think there is little selection also on the unobservables.

Finally, we have plotted the predicted probability based on the specification in the third column in Table 2, which conditions on only observable case characteristics on the right hand side of Figure 3. The observable case characteristics are seen to be unable to generate the cyclical patterns in the empirical frequencies compared to the homogenous specification on the left hand side, which conditions on the case characteristics and the extraneous factors.

	Homogenous with case characteristics	Homogeneous without case characteristics	Only case characteristics
intercept	−0.54 (0.91)	0.67 (0.23)**	−1.64 (0.84)
Extraneous factors			
pastparoles	0.36 (0.07)***	0.36 (0.07)***	
caseorder	−0.42 (0.07)***	−0.42 (0.06)***	
cumulativeminutes	0.00 (0.01)	0.00 (0.01)	
Case characteristics			
offence	0.05 (0.10)		0.11 (0.08)
imprison	−0.25 (0.06)***		−0.25 (0.05)***
monthsrv	−0.00 (0.00)		−0.00 (0.00)
rehab	1.96 (0.87)*		1.68 (0.76)*
nationality	−0.14 (0.15)		−0.05 (0.14)
gender	−0.30 (0.33)		−0.40 (0.28)
Judge specific leniency			
judge 2	−0.06 (0.24)	−0.09 (0.24)	−0.05 (0.20)
judge 3	−0.34 (0.28)	−0.37 (0.28)	−0.34 (0.28)
judge 4	−0.28 (0.24)	−0.33 (0.24)	−0.27 (0.23)
judge 5	−0.12 (0.28)	−0.15 (0.27)	0.14 (0.28)
judge 6	0.33 (0.29)	0.31 (0.31)	0.54 (0.30)
judge 7	0.03 (0.29)	0.02 (0.29)	0.10 (0.22)
judge 8	0.20 (0.28)	0.21 (0.29)	0.14 (0.24)
Num. obs.	1112	1112	1112

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Statistical models

Table 3: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
decision0	1,112	0.36	0.48	0	0	1	1
caseorder	1,112	5.40	3.85	1	2	7	25
cumulativeminutes	1,112	27.69	26.00	0	9	39	162
pastparoles	1,112	1.97	2.05	0	1	3	13
uniquesession	1,112	74.10	39.60	1	42	107.2	145
session	1,112	1.00	0.77	0	0	2	2
offence	1,112	2.83	0.94	0	2	3	7
imprison	1,112	1.94	1.52	0	1	2	13
monthsrv	1,112	28.92	34.21	0	12	30	300
nationlty	1,112	0.30	0.46	0	0	1	2
gender	1,112	0.95	0.22	0	1	1	1
rehab	1,112	0.98	0.13	0	1	1	1
judge.d1	1,112	0.21	0.41	0	0	0	1
judge.d2	1,112	0.20	0.40	0	0	0	1
judge.d3	1,112	0.08	0.26	0	0	0	1
judge.d4	1,112	0.15	0.35	0	0	0	1
judge.d5	1,112	0.07	0.25	0	0	0	1
judge.d6	1,112	0.05	0.22	0	0	0	1
judge.d7	1,112	0.14	0.35	0	0	0	1
judge.d8	1,112	0.11	0.31	0	0	0	1