THE CAUSAL EFFECT OF COGNITIVE ABILITIES ON ECONOMIC BEHAVIOR: EVIDENCE FROM A FORECASTING TASK WITH VARYING COGNITIVE LOAD

Ondřej Rydval

CERGE-EI

Charles University
Center for Economic Research and Graduate Education
Academy of Sciences of the Czech Republic
Economics Institute

WORKING PAPER SERIES (ISSN 1211-3298)
Electronic Version

457

Working Paper Series 457
(ISSN 1211-3298)

The Causal Effect of Cognitive Abilities on Economic Behavior: Evidence from a Forecasting Task with Varying Cognitive Load

Ondřej Rydval

CERGE-EI
Prague, April 2012

ISBN 978-80-7343-260-7 (Univerzita Karlova. Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-252-1 (Národohospodářský ústav AV ČR, v.v.i.)

The Causal Effect of Cognitive Abilities on Economic Behavior: Evidence from a Forecasting Task with Varying Cognitive Load[†]

Ondřej Rydval

Max Planck Institute of Economics*

CERGE-EI**

Abstract

We identify the causal effect of cognitive abilities on economic behavior in an experimental setting. Using a forecasting task with varying cognitive load, we identify the causal effect of working memory on subjects' forecasting performance, while also accounting for the effect of other cognitive, personality and demographic characteristics. Addressing the causality is important for understanding the nature of various decision-making errors, as well as for providing reliable policy implications in contexts such as student placement, personnel assignment, and public policy programs designed to augment abilities of the disadvantaged. We further argue that establishing the causality of cognitive abilities is a prerequisite for studying their interaction with financial incentives, with implications for the design of efficient incentive schemes.

Keywords: Cognitive ability; Causality; Experiment; Financial incentives; Performance; Working memory

JEL classification: C81, C91, D80, D83, J24

[†] I am grateful to Randall Engle and Richard Heitz (Attention and Working Memory Lab, Georgia Institute of Technology) for providing their "working memory span" tests, and to Brian MacWhinney (Department of Psychology, Carnegie Mellon University) for providing his "audio digit span" test. Andreas Ortmann and Nat Wilcox provided invaluable advice throughout the project. I further thank Colin Camerer, Dirk Engelman, Jan Hanousek, Glenn Harrison, Štěpán Jurajda, Jan Kmenta and Astrid Matthey for helpful comments. All errors remaining in this text are the responsibility of the author.

* Max Planck Institute of Economics*, Kahlaische Strasse 10, 07745 Jena, Germany.
Tel: +49-3641-686-641, Fax: +49-3641-686-667, E-mail: rydval@cerge-ei.cz

** Is a joint workplace of the Center for Economic Research and Graduate Education, Charles University, and the Economics Institute of Academy of Sciences of the Czech Republic.
Address: CERGE-EI, P.O. Box 882, Politických vězňů 7, Prague 1, 111 21, Czech Republic

Abstrakt

Tento článek se zabývá výzkumem kauzality kognitivních schopností v ekonomickém prostředí. V experimentální předpovídací úloze s proměnlivou kognitivní zátěží identifikujeme kauzální vliv operativní paměti (working memory) na výkonnost. Zároveň věnujeme pozornost vlivu dalších kognitivních, povahových a demografických charakteristik. Studium tohoto druhu kauzality je přínosné nejen pro pochopení různých chyb a odchylek v lidském chování, ale i pro politiku v oblastech jako je umísťování studentů do škol, výběr pracovníků ve firmách a veřejné programy zaměřené na zlepšování schopností znevýhodněných jedinců. Dále je identifikace této kauzality předpokladem pro porozumění současného vlivu kognitivních schopností a finančních odměn na výkonnost, z čehož plynou závěry pro tvorbu efektivních odměňovacích schémat.

1. Introduction

An extensive literature in economics and psychology has documented the predictive power of cognitive abilities for a variety of outcomes. Individuals with higher scores on various cognitive ability tests tend to behave closer to normative game-theoretic solutions (e.g., Burnham et al., 2009; Devetag and Warglien, 2003; Rydval et al., 2009); be less prone to behavioral biases and reasoning failures (e.g., Ballinger et al., 2011; Oechssler et al., 2009; Stanovich and West, 2000; Toplak et al., 2011); be less risk averse and time impatient (e.g., Benjamin et al., 2006; Dohmen et al., 2010; Frederick, 2005); and have more favorable socioeconomic outcomes (e.g., Cawley et al., 2001; Heckman et al., 2006). A parallel literature has shown that a similarly wide range of behavioral and socioeconomic outcomes correlate positively with favorable personality traits, most prominently with conscientiousness, locus of control and self-esteem (see, e.g., Almlund et al., 2011, and Ben-Ner et al., 2008, for reviews). However, with few exceptions (e.g., Ballinger et al., 2011; Heckman et al., 2006), the literatures have focused on measuring only few cognitive abilities *or* personality traits, rather than assessing the relative predictive power of a broader set of (theoretically relevant) abilities and traits.

More importantly, as argued in detail by Almlund et al. (2011), the majority of studies have only examined the predictive power of abilities and traits rather than their *causal* effect on outcomes. Interpreting the correlations is difficult since outcomes could predict measured individual characteristics or vice versa, and both could be caused by other factors. Addressing the reverse causality is important for understanding the nature of various decision-making errors, as well as for providing reliable policy implications in contexts such as student placement, personnel assignment, and public policy programs designed to augment abilities and traits of the disadvantaged. In the concluding section, we further argue that establishing the causality of individual characteristics is a prerequisite for credibly addressing issues relevant for the design of efficient incentive schemes, such as how people behave under different incentive levels and schemes conditional on their characteristics; how they self-select into the schemes based on their characteristics; and whether they are aware of their cognitive and personality limitations.

The causal effect of cognitive abilities and personality traits on socioeconomic outcomes has recently been addressed by labor economists using econometric modeling techniques. One strand of the literature uses structural equation methods that inevitably invoke debatable theoretical assumptions and identifying restrictions pertaining to the reverse causality running from outcomes to measured individual characteristics (e.g., Heckman et al., 2006). Another strand, with similar caveats, uses dynamic factor modeling to study the formation of abilities and traits over the life cycle (e.g., Cunha et al., 2010). Yet another strand is based on rare intervention studies aimed at improving specific individual characteristics of the disadvantaged, typically using random assignment to attribute the effect of a policy intervention on outcomes to the induced changes in the individual characteristics (e.g., Heckman et al., 2010).

In this paper, we demonstrate how one can identify the causal effect of *general* (or domain-general) cognitive abilities on economic behavior in a controlled experimental setting. Drawing on contemporary cognitive psychology, we broadly distinguish between general and task-specific cognitive abilities, and we choose general abilities to be represented by *working memory* – the capacity to control attention when executing cognitively complex tasks. Working memory tests are strong and robust predictors of general intelligence as well as performance in a broad range of tasks requiring maintenance of task-relevant information. Furthermore, compared to alternative measures of intelligence such as the Beta III test or the Raven test, working memory seems more firmly established theoretically, neurobiologically and psychometrically.

We identify the causal effect of working memory on performance in a time-series forecasting task consisting of a deterministic seasonal pattern “masked” by a state variable and an error term. Discovering the seasonal pattern and forecasting accurately requires maintaining forecast-relevant information accessible in memory while simultaneously processing it. Thus the task “activates” the type of cognitive ability that working memory theoretically represents. The causality test relies on manipulating the task's working memory load (or, more generally, the task's cognitive load). Two screens with forecast-relevant information are presented either simultaneously or sequentially. Since the sequential (simultaneous) presentation treatment features higher (lower) working memory load, working memory should, *ceteris paribus*, be a stronger (weaker) determinant of forecasting performance. We find this causality hypothesis confirmed for individual differences in asymptotic forecasting performance.

Ceteris paribus refers not only to the fact that other features of the forecasting task are identical for both treatments. It also means allowing for the possibility that, besides working memory, other cognitive abilities affect performance. We find that short-term memory – often regarded by cognitive psychologists as a task-specific cognitive ability counterpart of working memory – has a causal effect on performance parallel to that of working memory. On the other hand, basic arithmetic abilities – another task-specific ability – tend to predict performance only in the less memory-intensive simultaneous presentation treatment. Since other task-specific cognitive abilities such as prior forecasting expertise could be vital for performance but are hard to measure, we attempt to minimize their potential relevance by design and implementation features described in the next section. Nevertheless, we also obtain a proxy for prior forecasting expertise (or intrinsic forecasting ability) and show that controlling for it leaves our results intact.

We further account for individual heterogeneity in a broad set of personality traits that potentially affect forecasting performance. After controlling for the impact of the aforementioned cognitive abilities, performance is influenced negatively by math anxiety and, to a weaker extent, positively by *ex ante* intrinsic motivation. Other collected personality and demographic characteristics turn out irrelevant for performance. As a last performance determinant, we find that a subset of subjects who win a large windfall financial bonus immediately prior to the forecasting task are able to forecast better, but this effect occurs only in the more memory-intensive sequential presentation treatment.

We take further steps towards providing a clear interpretation of the causality of working memory and short-term memory. First, recent literature has argued that performance on cognitive tests is affected by various personality traits and economic preferences related mostly to test-taking motivation (e.g., Borghans et al., 2008). This issue might be relevant especially since, unlike the forecasting task, our cognitive tests were administered without performance-based financial incentives.[1] We find, for instance, that working memory scores are higher for subjects with higher perseverance and lower risk aversion. Parsing out these effects increases the predictive power of working memory in the sequential presentation treatment and provides stronger support for the causality hypothesis.

Second, working memory researchers often study the predictive power of the underlying working memory capacity to control attention, namely by removing the variance that working memory shares with short-term memory and other cognitive abilities (e.g., Engle et al., 1999). Following the practice, we remove the shared variance among working memory, short-term memory and basic arithmetic abilities. Doing so further increases the predictive power of working memory and short-term memory in the sequential presentation treatment and yields the strongest support for their causal effect among all model specifications. Importantly, this finding provides confidence that the high predictive power of working memory is not due to the shared surface features between the working memory test and the forecasting task (such as dealing with simple patterns and arithmetic operations), since the influence of the surface features was removed from the working memory scores when extracting the underlying capacity to control attention.

Our causality identification approach based on cognitive load manipulation has long been used to study the role of working memory in lower-order and higher-order cognitive processes.[2] Typical applications have been in the context of elementary attention tasks such as the antisaccade task (e.g., Kane et al., 2001) and the Stroop task (e.g., Kane and Engle, 2003). To our knowledge, we make the first attempt to employ the identification approach in the context of an economically relevant task. More

[1] We follow this standard psychology practice partly to avoid the possibility that any relationship found between the measured cognitive abilities and forecasting performance is due to them being incentivized in a similar manner.

[2] Alternatively, the literature has employed latent variable modeling, mainly to examine the relationship between working memory and general fluid intelligence (e.g., Kane et al., 2004).

generally, despite the wide-ranging predictive power of working memory in tasks studied by psychologists, working memory researchers themselves note almost complete lack of studies on the role of working memory in real-world problem-solving “insight” tasks requiring their solution to be gradually discovered (Hambrick and Engle, 2003).[3] Since many cognitively demanding, individual decision-making tasks in economics are “insight” tasks by their nature, we situate the causality test in such a setting.

Note that the identification approach examines whether increasing cognitive load strengthens the relationship between task performance and working memory, unlike the literature that studies the effect of increasing cognitive load on behavior itself (see, e.g., Duffy and Smith, 2011, for a review).[4] The latter literature can address neither the causality of cognitive abilities nor the role of cognitive abilities in general, at least not directly. Increasing cognitive load has been shown to produce, for instance, higher risk aversion (e.g., Benjamin et al., 2006), higher impulsiveness (e.g., Hinson et al., 2003) and less self control (e.g., Shiv and Fedorikhin, 1999), and hence these personality effects, rather than ability effects, could lie behind any observed treatment effect of cognitive load manipulation. We measure various traits related to impulsiveness and risk attitudes, but none of them turn out relevant for forecasting performance. Also, we find only a small and insignificant average treatment effect of the cognitive load manipulation on asymptotic forecasting performance, regardless of controlling for individual characteristics.

The paper is organized as follows. The next section describes the forecasting task and the causality identification approach. Section 3 reviews the measured cognitive, personality and demographic characteristics and summarizes the implementation details.

[3] As an exception, Welsh et al. (1999) find working memory to be correlated with performance in the Tower of London puzzle.

[4] Another difference between the two approaches is that the former usually manipulates cognitive load within the task itself (like we do), whereas the latter manipulates the load by distracting subjects with a secondary attention interference task. A related literature manipulates task complexity, usually making more extensive task adjustments compared to the cognitive load approaches. See, e.g., Bonner and Sprinkle (2002) for a review.

Section 4 presents the results, and Section 5 discusses their potential extensions and applications.

2. Experimental Design[5]

2.1 Forecasting task

We study individual behavior in a time-series forecasting task. Subjects repeatedly forecast a deterministic seasonal process, Ω_t, of the following form:

$$\Omega_t = B_t + \gamma_1 D_{1t} + \gamma_2 D_{2t} + \gamma_3 D_{3t} + \eta_t$$

$B_t \sim$ i.i.d. uniform {10, 20, 30, 40}
$D_{1t}=1$ if $t=1,4,7,\ldots 100$; 0 otherwise
$D_{2t}=1$ if $t=2,5,8,\ldots 98$; 0 otherwise
$D_{3t}=1$ if $t=3,6,9,\ldots 99$; 0 otherwise
$\gamma_1 = 46$, $\gamma_2 = 34$, $\gamma_3 = 18$
$\eta_t \sim$ *i.i.d.* uniform {-8, -4, 0, 4, 8}

Ω_t contains a state variable, B_t, a three-period seasonal pattern, $\Sigma_{s=1,2,3}\, \gamma_s D_{st}$, and an additive mean-zero *i.i.d.* error term, η_t. In each period *t*, subjects forecast the value of Ω_{t+1} based on observing eight-period "history windows," $(B_t,\ldots,B_{t-7})$ and $(\Omega_t,\ldots,\Omega_{t-7})$, on their screen. Subjects also observe B_{t+1} to be able to forecast Ω_{t+1}. They are told the distribution of B_t and η_t and about the existence of the seasonal pattern. Hence discovering the pattern and combining it with the observed values of B_{t+1} is the key to accurately forecasting Ω_{t+1}. After each forecast, F_{t+1}, subjects learn their current "noisy" forecast error, $\Omega_{t+1}-F_{t+1}$ (as opposed to the "true" forecast error, $\Omega_{t+1}-F_{t+1}-\eta_{t+1}$, the absolute value of which is used to measure forecasting performance).

B_t and $\Sigma_{s=1,2,3}\, \gamma_s D_{st}$ account for approximately equal shares of the total variance of Ω_t (namely 41% and 49%, respectively, with the remaining 10% attributable to η_t). As a consequence, the variability of B_t "masks" the seasonal pattern which cannot be inferred

[5] Additional design and implementation details are contained in the Appendix.

from past values of Ω_t alone but must rather be inferred from the differences between past values of Ω_t and B_t. Also, the presence of η_t means that subjects can only extract past values of Ω_t-$B_t = \gamma_s+\eta_t$. Hence discovering the seasonal parameters, γ_s, is a gradual, signal extraction task that likely taxes both working memory and short-term memory. The memory load does not cease entirely even after discovering the seasonal pattern since subjects continuously need to keep track of the revolving γ_s and to combine them with B_{t+1} in order to form their forecasts of Ω_{t+1}.

A sequence of pilots have indicated three key aspects of the cognitive complexity associated with extracting γ_s from $\gamma_s+\eta_t$ (henceforth Ω_t-complexity), namely, the number of values in the support of η_t; the degree of "overlap" of the $\gamma_s+\eta_t$ distributions, conditional on γ_s (i.e., their degree of non-monotonicity and non-uniqueness relative to each other); and the size of the "history window." Given the forecasting abilities in the student subject pool at hand, the present parameterization of γ_s and η_t has the convenient properties of bounding forecasting performance of a majority of subjects away from perfection throughout the task (and hence preserving financial incentives for learning) and generating sufficient potentially predictable between-subject variance in forecasting performance.

The character of the forecasting task reflects a consensus among psychologists on the cue-discovery nature of human learning in probabilistic environments. Even in the presence of random error, people seem proficient at discovering *which* cues are important (e.g., Dawes, 1979; Klayman 1984, 1988), as opposed to learning the exact weights attached to a given set of cues, especially correlated ones (e.g., Hammond et al., 1980; Brehmer, 1980). These findings have been largely confirmed by the time-series forecasting and expectation formation literatures: subjects are generally not good intuitive forecasters when it comes to determining parameter values of stochastic time series with even simple autoregressive or moving-average components (e.g., Hey, 1994; Maines and Hand, 1996); by contrast, subjects are good at detecting recognizable patterns in even relatively complex real-world time series (e.g., Lawrence and O'Connor, 2005). Thus our subjects should generally be capable of discovering the deterministic seasonal pattern even in the presence of randomness, η_t, but we challenge them further by introducing the state variable, B_t, that raises the memory load.

The time-series forecasting literature further documents that when the nature of the forecasted process permits so – for example, when the time series contains correlated past values or a trending component or both – subjects tend to employ various "natural" simplifying heuristics of the Kahneman and Tversky (1984) kind. They almost invariably anchor their forecasts on the most recent past value of the forecasted process and adjust it either for a previous trend (extrapolation heuristic), or for a long-term average (averaging heuristic), or for their previous forecast error(s) (exponential smoothing heuristic). These simplifying heuristics make forecasting strategies appear boundedly rational and ultimately reduce the overall memory load of forecasting tasks (e.g., Harvey et al., 1994; Hey, 1994). By choosing the forecasting process with a deterministic seasonal pattern, relatively high variance of B_t, and no trending component, we intentionally curb the effectiveness of the heuristics and create substantial opportunity cost to their use, as illustrated later for a specific averaging heuristic that we call a mechanical forecasting algorithm.

The fact that subjects know the distribution of B_t and η_t, combined with the detailed, example-oriented task instructions, make the forecasting task a logical rather than a statistical forward induction problem. This is meant to *a priori* minimize the influence of task-specific cognitive abilities that accrue from prior forecasting expertise. Prior expertise, or domain knowledge, is usually an important form of task-specific cognitive abilities,[6] but *individual differences* therein are hard to measure, so suppressing their potential importance seems desirable given our primary focus on the causal effect of general cognitive abilities. Another sense in which the impact of prior expertise is minimized is that forecasting performance is measured "asymptotically" after on-task learning has ceased.[7] It is nevertheless still possible that some form of "intrinsic

[6] See, e.g., Anderson (2000), Camerer and Hogarth (1999) and Libby and Luft (1993) for reviews. As noted, e.g., by Camerer and Hogarth (1999), prior expertise seems only imperfectly transferable across even slightly different cognitive production settings.

[7] Evidence from cognitive psychology suggests that on-task experience tends to be the most productive task-specific cognitive ability that often overrides the influence of prior expertise (e.g., Anderson, 2000;

forecasting ability" – such as pattern recognition skills in the presence of randomness – matters in our forecasting task, and that this ability is not well captured by the measured individual characteristics. In Section 4, we address this issue by obtaining a proxy for intrinsic forecasting ability. In Section 5, we propose how one could explicitly examine the effect of prior expertise.

2.2 Causality identification approach

The experimental design consists of two between-subject treatments that vary in their working memory load, and likely also in their short-term memory load. In the treatment with higher memory load, T_{seq}, the two screens with the values of $(B_{t+1},\ldots,B_{t-7})$ and $(\Omega_{t},\ldots,\Omega_{t-7})$, respectively, are in each period *t* displayed sequentially. By contrast, in the treatment with lower memory load, T_{sim}, the two screens are displayed simultaneously.[8] The treatment variation permits identifying the causal effect of working memory on forecasting performance by testing the following causality hypothesis:

> Hypothesis: *Ceteris paribus*, since T_{seq} features higher working memory load compared to T_{sim}, working memory has a stronger positive impact on forecasting performance in T_{seq} compared to T_{sim}.

To see the difference in the memory load between T_{seq} and T_{sim}, recall that in order to extract the seasonal pattern, subjects need to attend to the differences between past values of Ω_t and B_t. Doing so is more memory-intensive in T_{seq} where past Ω_t-B_t values must be calculated virtually, leaving less scarce memory resources for extracting the seasonal pattern. By contrast, subjects in T_{sim} can calculate past Ω_t-B_t values visually from the simultaneously presented $(B_{t+1},\ldots,B_{t-7})$ and $(\Omega_{t},\ldots,\Omega_{t-7})$ screens. Hence T_{sim} supplies "external memory" for the calculation of past Ω_t-B_t values, which relaxes the

---

Ericsson and Smith, 1991; Reber, 1989). Findings from experimental economics seem less conclusive (e.g., Kagel and Levine, 1986, and the ensuing debate).

[8] In T_{sim}, subjects observe the two parallel screens for 15 seconds. In T_{seq}, subjects observe the $(B_{t+1},\ldots,B_{t-7})$ screen for 10 seconds and subsequently the $(\Omega_{t},\ldots,\Omega_{t-7})$ screen for 15 seconds. The working memory literature illustrates that (sensible) time constraints, and, more generally, individual differences in effort duration, are inconsequential for the relationship between working memory and cognitive performance (e.g., Engle and Kane, 2004; Heitz et al., 2008).

memory load of the calculation and leaves more memory resources for the actual extraction of the seasonal pattern.

The cognitive load imposed in T_{seq} closely matches the aspects of cognition theoretically underlying the working memory construct, namely maintenance of relevant information in active memory, resolution of conflicting information, and controlled allocation of attention (Engle and Kane, 2004). Forecasting in T_{seq} predominantly requires the use of System 2 (controlled processing) type of cognitive ability, of which working memory is a fundamental component. On the other hand, forecasting in T_{sim} is likely to pose a much more reflexive, pattern-recognition exercise requiring mostly the use of System 1 (automated processing) type of cognitive ability (e.g., Feldman-Barrett et al., 2004; Stanovich and West, 2000).

2.3 Forecasting sequences and payoff function

The forecasting sequences, Ω_t, vary across subjects but are not generated completely at random. In order to retain basic control over the influence of Ω_t-complexity on between-subject variance in performance, only the η_t streams vary across subjects,[9] and these are representative in terms of several theoretically relevant aspects of Ω_t-complexity (see the Appendix for details). Further, to obtain "cleaner" across-treatment comparisons of performance and its determinants, we use the same set of Ω_t sequences in both treatments. As detailed in Section 4.3, we then remove the impact of Ω_t-complexity by comparing performance of the pairs of subjects facing identical Ω_t sequences across treatments.

We measure performance in a couple of twelve-period segments of the forecasting task, namely in the EARLY periods 21-32 and the LATE periods 84-95. For each subject, the EARLY and LATE segments of Ω_t as well as the eight periods preceding them are exactly matched in terms of all the Ω_t components on a period-by-period basis. Each subject thus forecasts the same segment of his or her Ω_t sequence twice, based on

[9] For all subjects, B_t consists of the same sequence of permutations on the support of B_t. The permutations are selected and adjoined in such a way as to avoid repeating values and easily memorable sequences. Further, each B_t value is paired with each value of the seasonal pattern in approximately equal frequencies.

observing the same forecast-relevant information.[10] One advantage of this design feature is that a comparison of each subject's EARLY and LATE performance yields an unambiguous within-subject measure of learning. Another advantage is that the correlation between EARLY and LATE performance provides a useful indicator of the internal reliability of the chosen forecasting performance measure. Finally, we will argue that residual variation in EARLY performance (after parsing out the influence of individual characteristic) serves as an efficient proxy for prior forecasting expertise in the estimated model of LATE performance.

The payoff function has the form of a betting scheme. At the very beginning of each period, i.e., prior to observing the screens with forecast-relevant information, subjects are asked to bet an amount x_t on their forecast, F_{t+1}. They can bet up to 100 ECU, but at least 50 ECU so that they always have sufficient financial incentives to forecast accurately. The payoff (in ECU) in period t, π_t, then depends on the "noisy" absolute forecast error, $abs(\Omega_{t+1}\text{-}F_{t+1})$, as well as on the amount bet, x_t:

$$\pi_t = x_t\theta g_t + (1\text{-}\theta)(100\text{-}x_t)$$

where $g_t = max\{20 - abs(\Omega_{t+1}\text{-}F_{t+1}),0\}$, $50 \leq x_t \leq 100$, and θ=0.1

Hence the return to betting, θg_t, is a negative linear function of the "noisy" absolute forecast error, as long as the forecast error does not exceed 20 whereby the return to betting becomes zero. On the other hand, every ECU not bet earns a riskless return of $(1-\theta)$. Clearly, betting $x_t>50$ is profitable only if $g_t>(1\text{-}\theta)/\theta$, i.e., only if $abs(\Omega_{t+1}\text{-}F_{t+1})<11$. The net gain from betting $x_t>50$ hence becomes positive only if subjects manage to reduce their "noisy" absolute forecast errors below 11 on average.

[10] Reflecting findings from pilots, EARLY performance is measured sufficiently after the beginning of the forecasting task to ensure task salience. LATE performance is measured as much apart from EARLY performance as possible, but sufficiently away from the end of the task to avoid lapses of concentration affecting the performance measure. It is in our view unlikely (and there is no evidence in the debriefing questionnaire) that subjects would recognize the repeated part of the Ω_t sequence after more than 60 periods. This is especially due to the stationary nature of Ω_t, implying that the EARLY and LATE segments do not differ in an easily recognizable manner from other Ω_t segments. Nevertheless, we checked that when performance is instead measured in the twelve periods directly preceding the LATE segment, the results of the multivariate analysis presented in Section 4.3 qualitatively hold.

The parameterization of the payoff function is conveniently linked with the parameterization of the Ω_t process. To see this, consider forecasting performance of a mechanical forecasting algorithm that, instead of focusing on extracting the seasonal pattern, forms its point forecast simply by adding B_{t+1} to the average of the three most recent past values of Ω_t-B_t. The mean "noisy" absolute forecast error of the algorithm is about 11.3 on average (varying slightly across different Ω_t sequences; the mean true absolute forecast error is about 10.3). Hence to find betting x_t>50 profitable, subjects must perform better than the mechanical forecasting algorithm: they must attempt to discover the seasonal pattern.

One reason we make subjects bet on their forecasts is to keep the relatively lengthy forecasting task intellectually stimulating throughout. As discussed in the concluding section, another reason is to obtain a decision-relevant, incentive-compatible measure of confidence in forecasting abilities, and to analyze how this confidence influences forecasting performance. Although analysis of betting behavior is not the subject of this paper, we show that subjects' incentives to forecast accurately were comparable across treatments. Namely, in the LATE segment for which we mostly analyze performance, the average bet is 85.8 ECU and is only marginally higher in T_{sim} by 6% (p=0.097). In fact, 44% of subjects in T_{sim} and 35% in T_{seq} bet the maximum of 100ECU throughout the LATE segment, and only 15% in T_{sim} and 25% in T_{seq} bet less than 70 ECU on average. In the EARLY segment, subjects bet considerably less, namely 68.4 ECU on average, and their bets are on average higher in T_{sim} by a larger margin of 14% (p=0.002).[11] Still, all subjects of course bet at least the required 50 ECU.

3. Measured Individual Characteristics and Implementation Details

3.1 Cognitive abilities

We measure *working memory* by a working memory span test, namely a computerized version of the "operation span" test (Turner and Engle, 1989). The test requires memorizing sequences (of various lengths) of briefly presented letters interrupted by

[11] The across-treatment comparison of bets is performed using the paired two-tail Wilcoxon signed-rank test, pairing subjects with identical Ω_t sequences in T_{sim} and T_{seq}.

solving simple arithmetic problems (i.e., by an attention interference task).[12] At the end of a sequence, subjects recall as many letters as possible in the correct positions in the sequence. The working memory score is the number of correctly recalled letters summed across all sequences.[13]

Working memory constitutes theoretically and neurobiologically a well-defined domain-general cognitive ability, representing the capacity to control attention (e.g., Engle and Kane, 2004). Working memory span tests have strong internal reliability (e.g., Conway et al., 2005). They predict performance in general fluid intelligence tests as well as in a broad range of lower- and higher-order cognitive tasks requiring controlled (as opposed to automated) information processing, such as reading comprehension, abstract reasoning, problem solving, and complex learning (see, e.g., Feldman-Barrett et al., 2004). Ackerman et al. (2002) suggest that working memory span tests are both theoretically and psychometrically superior to alternative, potentially broader tests of general intelligence such as the Beta III test (Kellogg and Morten, 1999) or the Raven test (Raven et al., 1998).[14] Kane et al. (2004) argue that, in trying to understand the effect of general cognitive abilities on behavior, one ought to start with exploring rather reductionist measures such as working memory, preferring clarity of interpretation over breadth of measurement. We follow this approach.

We measure *short-term memory* by a computerized auditory "digit span" test, similar to the Wechsler digit span test (e.g., Devetag and Warglien, 2003). The test requires memorizing pseudo-random sequences (of various lengths) of briefly presented digits and recalling them in the correct position in the sequence. The short-term memory score is the number of recalled digits summed across all sequences.[15]

---

[12] Subjects determine, in a true-false manner, whether equations such as "(9/3)-2=2?" are solved correctly. In an initial practice period, the computer measures each subject's equation-solving speed and subsequently requires the subject to maintain the speed throughout the test while also maintaining solution accuracy.

[13] Alternative scoring procedures are described in Conway et al. (2005).

[14] In Ballinger et al. (2011), both working memory and Beta III (namely its two analytical components) are positively correlated with performance in a precautionary saving task.

[15] From several alternative scoring procedures, we use the one that is most directly comparable to the working memory test scoring procedure.

Short-term memory (simple) span tests such as ours are thought to reflect information storage capacity as well as information coding and rehearsal skills that make the stored information more memorable (e.g., Engle et al., 1999). Such coding and rehearsal strategies are assumed to be eliminated from working memory span tests through the presence of an attention interference task, which in turn is the only differentiating design feature ensuring that the working and short-term memory span tests measure separate cognitive constructs. In our forecasting setting, better short-term memory might, for instance, increase the number of past Ω_t-B_t values that subjects are able to calculate, especially in T_{seq} where past Ω_t and B_t values appear on separate screens. The working memory literature extensively documents that short-term memory is a more task-specific (rather than domain-general) cognitive ability, in that it is not as strongly related to general intelligence and to performance in tasks requiring controlled information processing as is working memory.[16] In fact, the literature usually views working memory and short-term memory as comprising a functional working memory *system*, with working memory being the central component representing the capacity to control attention and short-term memory being the supporting storage, coding and rehearsal component (e.g., Kane et al., 2004; Heitz et al., 2005).

As another potentially relevant cognitive ability, we measure basic *math* abilities under time pressure. We administer an "addition and subtraction" test with 60 items and a two-minute time limit. The test sheet has alternating rows of 2-digit additions and subtractions, such as "25+29=__" or "96–24=__." The math score is the total count of correct answers.[17]

Our math test belongs to the class of basic arithmetic ability tests provided by the "ETS Kit of Referenced Tests for Cognitive Factors" (Ekstrom et al., 1976). The tests are assumed to measure the ability to perform basic arithmetic operations with speed and accuracy but are not meant to capture mathematical reasoning or higher mathematical

[16] This is particularly true if short-term memory is measured by verbal or numerical tests. Spatial tests have more general predictive power (e.g., Kane et al., 2004).

[17] The test in fact had two parts containing different items, separated by a couple of other tasks that took 15-20 minutes. The test-retest reliability, as measured by the Pearson correlation coefficient between the two test scores, is 0.854.

skills. The test closely matches the basic arithmetic skills required in the forecasting task and hence can be regarded as another task-specific cognitive ability measure. While we have no strong priors as regards the relative impact of basic arithmetic abilities on forecasting performance across treatments, the impact is likely to be overridden by working and short-term memory requirements of the more memory-intensive sequential presentation treatment.[18]

3.2 Personality traits and demographic characteristics

The personality traits described below are measured as potential correlates of forecasting performance and cognitive abilities. Also, as mentioned below, some of the personality traits correlate with each other, so measuring them separately seems desirable for disentangling their impact. Each personality trait is measured by a ”personality scale” consisting of a collection of 10-12 statements. Subjects indicate their agreement or disagreement with each of the statements as follows: 1 = “entirely true,” 2 = “mostly true,” 3 = “mostly false” and 4 = “entirely false.” Subjects are told that there are “neither good nor bad choices” and are asked to make choices most closely reflecting their attitudes and behavior. Since both positively and negatively worded statements are included, the choices for negatively worded statements are recoded (with one exception noted below). Each subject’s score is the average of his or her choices. All the personality scales are included in a single sheet and subjects encounter the various statements in a randomized order identical across subjects.

We measure intrinsic motivation by a scale called *need for cognition*, a well-established measure of one’s motivation to engage in effortful, cognitively demanding tasks (e.g., Cacioppo et al., 1996).[19] Need for cognition is closely connected with conscientiousness – a personality trait from the well-known “Big Five” trait inventory – which has been

[18] One may further wish to measure perceptual speed abilities as these apparently matter for basic encoding and comparison of items under time pressure (e.g., Ackerman et al., 2002). Nevertheless, complex perceptual speed test scores and working memory span test scores share substantial variance, and the causality appears to run from working memory to perceptual abilities rather than vice versa (e.g., Heitz et al., 2005).

[19] As in Ballinger et al. (2011), we use a short version of the need for cognition scale of Cacioppo et al. (1984). Ballinger et al. (2011) find need for cognition to be positively correlated with performance in their precautionary saving task, but the relationship vanishes once the impact of cognitive abilities is taken into account.

documented as the most prominent predictor of behavioral and socioeconomic outcomes (e.g., Almlund et al., 2011).[20] There is an extensive but inconclusive literature in economics, psychology and neuroscience on the channels through which financial incentives could interact with (e.g., crowd-out) intrinsic motivation in stimulating effort and performance (e.g., Deci et al., 1999; Eisenberger and Cameron, 1996; Gneezy and Rustichini, 2000; McDaniel and Rutström, 2001; Murayama et al., 2010). Not directly addressing the complex interactions, we include intrinsic motivation in the empirical model of forecasting performance in a reduced-form manner to account for the possibility that subjects are *ex ante* differentially motivated to perform well, especially in the more cognitively demanding T_{seq} treatment.

The *math anxiety* scale (e.g., Pajares and Urdan, 1996) measures anxiety or feelings of tension when manipulating numbers or solving math problems. For this scale only, we recode the choices for *positively* worded statements so that a high score means high math anxiety. Anxiety to deal with numbers (under time pressure) could affect forecasting performance as well as performance on the cognitive tests. In fact, math anxiety has been found correlated with mathematics achievement, aptitude and schooling grades (e.g., Pajares and Miller, 1994; Schwarzer et al., 1989), and it is closely related to other math-related psychological constructs such as math self-efficacy and math self-concept (e.g., Cooper and Robinson, 1991; Pajares and Miller, 1994).

We also use three scales of Whiteside and Lynam (2001) that are meant to measure various aspects of impulsive behavior.[21] Impulsiveness might matter in our context since Hinson et al. (2003) find it to be more prominent under higher cognitive load. More specifically, the *sensation-seeking* scale is a broad proxy for risk-taking attitude, which might affect subjects' willingness to experiment with alternative forecasting strategies, for instance with alternative approaches to discovering the seasonal pattern. Sensation-seeking has been found positively correlated with need for cognition (e.g., Crowley and Hoyer, 1989) and risk-taking behavior (e.g., Eckel and Wilson, 2004). The

[20] Conscientiousness is viewed as the tendency to be organized, responsible, dependable, hardworking and persistent.
[21] Ballinger et al. (2011) find that, of the three impulsiveness scales, only perseverance predicts performance in a precautionary saving task, but its effect is unexpectedly negative.

premeditation scale is a proxy for one's propensity to pause and think carefully while carrying out (cognitive) tasks, which might be relevant for forming successful forecasting strategies, possibly complementing sensation-seeking. Lastly, the *perseverance* scale is thought to measures one's determination in solving lengthy and demanding tasks, which might matter since our key measure of forecasting performance is situated towards the end of the rather lengthy task.

We further measure *judgmental confidence*, or trust in one's judgment, by a short version of the Judgmental Self-Doubt Scale (Mirels et al., 2002). The authors show that the scale is positively correlated, for instance, with need for cognition, as well as with self-esteem and locus of control which both have been found associated with a variety of behavioral and socioeconomic outcomes (see, e.g., Almlund et al., 2011). Higher judgmental confidence is also associated with lower anxiety, causal uncertainty, discomfort with ambiguity and preference for predictability, all of which could affect forecasting performance given the stochastic features of the task. We also measure judgmental confidence as a likely determinant of betting behavior.

In addition to the above personality scales, we also measure subjects' *risk aversion* using a hypothetical risk elicitation task of the Holt and Laury (2002) format.[22] While risk aversion should not theoretically influence forecasting decisions *per se* since they are risk-free in the economic sense, risk attitudes could affect the formation of forecasting strategies, as hypothesized above for sensation-seeking. Especially if sensation-seeking turns out important for forecasting behavior, one may wish to have a proxy for risk attitudes as usually measured by economists. Furthermore, there is also growing evidence that people with higher cognitive abilities tend to be less risk averse (e.g., Dohmen et al., 2010), so risk aversion could be negatively correlated especially with working memory.

[22] The risk elicitation task has six "multiple price lists," each consisting of 11 ordered risky choice pairs. Subjects draw a horizontal line to indicate their willingness to switch from a fixed sure payoff to an increasingly attractive gamble. The risk aversion score is constructed as the summation of the line locations. The risk elicitation task in fact had two identical parts separated by a couple of unrelated tasks that took 15-20 minutes. The test-retest reliability of the risk aversion measure, as indicated by the Pearson correlation coefficient between the two test scores, is 0.932. Although the results of hypothetical and incentivized risk elicitation may differ (e.g., Harrison and Rutström 2008), we prefer to collect all the cognitive and personality characteristics in the same, non-incentivized manner.

In a short questionnaire, we further collected demographic variables, namely age, gender and university field of study. The questionnaire also collected proxies for socioeconomic status, namely the number of household members and functional cars in the household in the last year of high school, and a binary indicator of one's current personal car ownership. None of the socioeconomic variables turn out important in the analysis below.

Lastly, right after completing the collection of individual characteristics, subjects had a chance to win a substantial financial bonus, later referred to as *windfall* because it was awarded exogenously with respect to the forecasting task.[23] The bonus affected 14 (out of 124) subjects and amounted to 750CZK for 13 subjects and 1500CZK (approximately PPP$117) for the remaining one. The incidence of the bonus was balanced across treatments: eight bonuses were awarded to subjects in T_{seq}. We had no strong priors as to whether the bonus ought to foster or discourage *ex ante* intrinsic motivation to forecast well, and how the bonus interacts with the high-powered financial incentives implemented in the forecasting task itself. The bonus could induce positive affect, which has been shown to foster flexible thinking and problem solving and enhance performance even in difficult, complex tasks (see, e.g., Isen, 2000, for a review). Isen and Reeve (2006) further argue that positive affect fosters intrinsic motivation, performance on enjoyable tasks, and responsible work behavior on uninteresting tasks.

3.3 Implementation details

The experiment was conducted at the Bank Austria Portable Experimental Laboratory at CERGE-EI, Prague. The subjects were 124 full-time native Czech students recruited from Prague universities and colleges.[24] The treatments were balanced in terms of gender (32 and 34 males in T_{sim} and T_{seq}, respectively).

[23] In each experimental session, we conducted a short guessing game experiment from which 2-3 randomly selected subjects won the bonus, the size of which depended on their choice in the game and the number of winners. The possibility of winning the bonus was pre-announced in the initial instructions.
[24] Further information on the subject pool and recruitment procedures is contained in the Appendix.

Experimental sessions lasted approximately 4 hours on average, but no longer than 4.5 hours. The collection of individual characteristics in the first part of each session usually lasted 1.5-2 hours. The order of collection was the same across sessions, with the cognitive tests generally preceding the personality scales and the demographic questionnaire. For the completion, subjects earned a participation fee of 150 CZK (approximately PPP$12) and could win the windfall financial bonus. The working memory and short-term memory tests were conducted using E-prime (Schneider et al., 2002) while the remaining collection was administered in a paper-and-pencil format.

After a short break, the forecasting task, programmed and conducted in z-Tree (Fischbacher, 2007), lasted about two hours and was completed at each subject's individual pace. All periods were paid and subjects could earn over 900CZK (approximately PPP$70). The actual average earnings across both treatments were 483CZK (approximately PPP$38). Thus together with the participation fee, subjects walked out with PPP$50 on average (not counting the windfall financial bonus). After completing a debriefing questionnaire, subjects were paid off privately in cash. All parts of the experiment were conducted anonymously.

4. Results

4.1 <u>Forecasting performance</u>

As mentioned earlier, subject *i*'s forecasting performance in period *t* is measured in terms of his or her "true" absolute forecast error, $abs(\Omega_{i,t+1}\text{-}F_{i,t+1}\text{-}\eta_{i,t+1})$, henceforth "forecast error" unless otherwise noted. Let $M_{i,t}$ denote subject *i*'s twelve-period moving average of forecast errors up to period *t*. Figure 1 displays the evolution of average $M_{i,t}$ in each treatment. Average performance is clearly better in the less memory-intensive simultaneous presentation treatment, T_{sim}. The difference relative to T_{seq} – and thus the average treatment effect of the memory load manipulation – becomes smaller over time. There is a considerable extent of learning, especially in the first half of the task. In both treatments, average $M_{i,t}$ about halves throughout the task. The evolution of performance can be judged relative to the benchmark of the above mentioned mechanical forecasting algorithm with the mean true absolute forecast error of about 10.3. In both treatments, average $M_{i,t}$ gradually falls below the benchmark, though more than twice faster in T_{sim}. In T_{seq}, average $M_{i,t}$ starts much further above the benchmark and takes around 40 forecasting periods to reach it.[25]

Since the forthcoming analysis focuses on explaining performance heterogeneity, Figure 1 shows the 10th and 90th percentiles of $M_{i,t}$ to illustrate that both treatments generate plenty of potentially predictable between-subject variance throughout the task. The worst forecasters perform similarly across treatments, whereas the best forecasters consistently perform better in the less memory-intensive T_{sim}. In both treatments, even the worst forecasters show some learning progress, and even the best forecasters always have financial incentive to improve performance and do so. As an exception, a small fraction of subjects in both treatments reach the performance ceiling (slightly more often in T_{sim}), which might reduce the extent of between-subject performance variability and attenuate the predictive power of individual characteristics. This concern is addressed in the multivariate analysis below and turns out to be of minor importance.

An additional source of performance heterogeneity not apparent from Figure 1 is the seasonal nature of the forecasting task. Performance varies across the three forecasting seasons, with the "sandwich" seasonal parameter, $\gamma_2 = 34$, usually being associated with markedly lower and less variable forecast errors. The forecasting seasons feature various degrees of "overlap" of the $\gamma_s+\eta_t$ distributions, conditional on γ_s, which seems to affect the relative ease of discovering the seasonal parameters, γ_s. While a more detailed seasonal analysis is possible, a potential caveat is that different (unobserved) forecasting strategies may imply different seasonal performance tradeoffs, which may in turn limit interpretability of the results. We prefer to adopt a more conservative approach by aggregating performance across seasons.

To look closer at the across-treatment differentials in performance as well as the extent of learning, we focus on performance in the perfectly matched EARLY and LATE twelve-period segments, namely $M_{i,31}\equiv M_{i,\text{EARLY}}$ and $M_{i,94}\equiv M_{i,\text{LATE}}$, respectively (we drop the *i* subscript whenever the context is clear). Confirming the general observations

[25] Recall that subjects make their first forecast in period 8.

from Figure 1, the average M_{EARLY} is significantly lower in T_{sim} at 10.04 compared to 13.58 in T_{seq} (p=0.002).[26] The average M_{LATE} is also lower in T_{sim} at 5.81 compared to 6.93 in T_{seq}, but the difference is insignificant (p=0.249). Learning, measured as $M_{i,EARLY}$-$M_{i,LATE}$, is considerable and significant at p<0.001 in either treatment, but its extent is significantly greater in T_{seq} (p=0.020), mainly due to the higher M_{EARLY} in that treatment.[27]

Despite the considerable distance between their measurement, M_{EARLY} and M_{LATE} are strongly correlated with each other, the Spearman rank correlations reaching 0.726 (p<0.001) in T_{sim} and 0.450 (p<0.001) in T_{seq} (the corresponding Pearson correlations are almost identical). Especially the former correlation suggests strong internal reliability of our measure of forecasting performance, with implications for testing the causality hypothesis. Namely, if working memory indeed turns out to be a stronger predictor of performance in T_{seq} compared to T_{sim}, this is unlikely due to lower internal reliability of the performance measure in T_{sim} but rather due to the causal effect of working memory on performance.

In testing the causality hypothesis, we focus on M_{LATE}, which can be regarded as "asymptotic" performance in the sense that learning has ceased. In particular, $M_{i,LATE}$ does not differ from performance in the previous twelve-period segment, $M_{i,82}$, in either treatment (p=0.392 in T_{sim}, p=0.997 in T_{seq}).[28] As a robust alternative to M_{LATE}, free of occasional performance "slip-ups," we further consider another performance measure based on the average of seasonal medians of forecast errors in the LATE segment. Since

[26] All statistical comparisons in this section are conducted using the paired two-tail Wilcoxon signed-rank test. Across-treatment comparisons pair subjects with identical Ω_t sequences in T_{sim} and T_{seq}. The results are qualitatively confirmed by the paired t-test as well as corresponding unpaired tests.

[27] To get a perspective on the extent of learning, note that the average $M_{i,EARLY}$-$M_{i,LATE}$ (4.24 in T_{sim}, 6.65 in T_{seq}) is comparable in magnitude to its standard deviation (4.18 in T_{sim}, 5.59 in T_{seq}) as well as the standard deviation of M_{EARLY} (5.87 in T_{sim}, 5.54 in T_{seq}) and M_{LATE} (5.36 in T_{sim}, 5.43 in T_{seq}) themselves.

[28] On the other hand, $M_{i,82}$ significantly differs from $M_{i,70}$ in both treatments (p=0.032 in T_{sim}, p<0.001 in T_{seq}).

the two measures are almost perfectly correlated and the results are nearly identical for both of them, we present the results only for M_{LATE}.[29]

4.2 Bivariate relationships between performance and individual characteristics

We first briefly summarize the collected cognitive, personality and demographic characteristics. None of them differ significantly across treatments at $p<0.1$ using the unpaired two-tail Wilcoxon ranksum test and *t*-test.[30] As detailed in Sections 3.2 and 3.3, the treatments are, for instance, balanced in terms of gender and the incidence of the windfall bonus. All the characteristics have considerable variability that does not differ significantly across treatments at $p<0.1$ using the two-tail *F*-test.[31]

Table 1 displays Spearman rank correlations between subjects' forecasting performance and their cognitive, personality and demographic characteristics for T_{sim} and T_{seq}. The correlation between M_{LATE} and Working memory is high and significant at -0.479 in T_{seq}, compared to the much lower and insignificant correlation of -0.109 in T_{sim}. Hence in line with the causality hypothesis, higher working memory is more strongly associated with better performance when the memory load is higher. In the next section, we confirm this conclusion when other potential predictors of M_{LATE} are taken into account. To that end, Table 1 shows that M_{LATE} in T_{seq} is also negatively correlated with Short-term memory and Math, though only about half the magnitude compared to the correlation of M_{LATE} with Working memory. M_{LATE} also improves with higher Premeditation and lower Math anxiety in T_{sim}, and is better for males in T_{seq}. However, the gender effect vanishes in the multivariate analysis.

Building on recent literature (e.g., Almlund et al., 2011; Borghans et al., 2008; Segal, 2008), we next parse out the influence of personality traits from the cognitive test

---

[29] The same holds for an analogously constructed robust alternative to M_{EARLY}. One might actually not wish to remove performance slip-ups if they arise from momentary distraction related to individual differences in working memory.

[30] As an exception, Working memory is marginally higher in T_{sim} by the ranksum test (p=0.081) but not by the *t*-test (p=0.458).

[31] As an exception, Short-term memory is more variable in T_{sim} ($p<0.001$) due to an outlier at the bottom performance end, the exclusion of which has no effect on any of the reported results. Potentially slightly attenuating the predictive power of individual characteristics, the maximum score is reached by four subjects for Working memory (two per treatment), one for Sensation-seeking, and one for Math anxiety. The minimum score is reached by one subject for Risk aversion and four for Math anxiety.

scores. Working memory is higher for subjects with higher Perseverance and lower Risk aversion. The partial correlations – i.e., standardized regression coefficients from a OLS regression of Working memory on Perseverance and Risk aversion in the pooled sample – are 0.189 (p=0.048) and -0.122 (p=0.104), respectively.[32] Perseverance could capture certain aspects of test-taking motivation,[33] while the Risk aversion correlation is in line with the aforementioned literature documenting that more intelligent people tend to be less risk averse. We call the residuals from the above regression WMresid1. Table 1 shows that, in T_{seq}, WMresid1 correlates with M_{LATE} as strongly as does Working memory itself; the corresponding correlation in T_{sim} is now even smaller.

Analogous to above, we find that Short-term memory is not significantly related at p<0.1 to any personality or demographic variables. On the other hand, Math is higher for males with lower Risk aversion and females with lower Math anxiety; the partial correlations obtained from a OLS regression accounting for these gender interactions are -0.330 (p=0.006) and -0.367 (p<0.001), respectively.[34] While the impact of Math anxiety is anticipated, it is interesting to establish it only for females. The Risk aversion correlation resembles the finding of Dave et al. (2010) obtained for a broader measure of mathematical abilities and for both genders.[35] We call the residuals from the above regression MATHresid1. Table 1 shows that MATHresid1 correlates with M_{LATE} slightly less than does Math.

[32] Although the Working memory score reaches the maximum for two subjects in each treatment, we report OLS results as these are virtually identical to those obtained from a censored-type estimation. In the estimations reported in this section, p-values are based on heteroskedasticity-robust standard errors. Personality and demographic variables that are not included in the estimations (as well as any interactions and higher-order moments) are individually and jointly highly insignificant. The reported correlations are Spearman rank correlations.

[33] Need for cognition turns out highly insignificant, broadly in line with evidence suggesting that working memory scores are not correlated with cognitive effort exerted during the tests (e.g., Heitz et al., 2008). Nevertheless, Need for cognition is highly correlated with both Perseverance (0.237, p=0.008) and Risk aversion (-0.301, p<0.001).

[34] The listed coefficients pertain to males and females, respectively, while the corresponding coefficients for the other gender are close to zero and highly insignificant. Both the gender interactions are significant at p<0.05, whereas the gender dummy is highly insignificant and hence omitted. Math is also significantly lower at p<0.1 for older subjects, but Age is not included in the final estimation as we see no theoretical justification for parsing it out from Math.

[35] Both Risk aversion and Math anxiety are negatively correlated with Need for cognition (at p<0.01) and Perseverance (at p<0.1), both of which proxy for motivation but are insignificant when included in the estimation.

To examine the separate predictive power of the underlying working memory capacity to control attention, we extend the above Working memory estimation by parsing out the shared variance with Short-term memory and Math, as well as any statistically significant personality variables.[36] Namely, the partial correlation coefficients are 0.399 (p<0.001) for Short-term memory and 0.133 (p=0.099) for Math, while also controlling for the positive effect of Perseverance and negative effects of Risk aversion and Need for cognition.[37] We call the residuals from the above regression WMresid2. Table 1 shows that, relative to Working memory and WMresid1, WMresid2 correlates with M_{LATE} less strongly but still highly significantly in T_{seq}; the corresponding correlation in T_{sim} is again negligible.

Analogously, we further parse out the effects of Working memory, Premeditation and Judgmental confidence from the Short-term memory score.[38] Table 1 shows that the regression residuals, STMresid2, are uncorrelated with M_{LATE} in either treatment. This suggests that the significant correlation between Short-term memory and M_{LATE} in T_{seq} was due to the shared variance between Short-term memory and Working memory. Lastly, we parse out the effects of Working memory, Risk aversion, Math anxiety and Sensation-seeking from the Math score.[39] Table 1 shows that the regression residuals, MATHresid2, are uncorrelated with M_{LATE} in either treatment, which again suggests that the correlation between Math and M_{LATE} in T_{seq} was due to the shared variance

[36] Approaches to extracting the capacity to control attention vary. Depending on the research goal, either the shared or the residual variance between working memory and short-term memory is parsed out (e.g., Engle et al., 1999; Kane et al., 2004). We adopt an approach that most closely corresponds to the way in which we parse out personality variables.

[37] Short-term memory and Math together explain 20% of the variance in Working memory while the personality variables explain additional 6 percentage points. The negative effect of Need for cognition could be compensation for the fact that Short-term memory or Math to some extent proxy for test-taking motivation as they are measured without performance-based incentives. Note that Need for cognition is uncorrelated with Working memory itself as well as with how subjects perform on the working memory test's equation-solving speed test (see Section 3.1).

[38] Working memory explains 17.5% of the variance in Short-term memory, while the personality variables explain additional 2.5 percentage points. Math is highly insignificant. The effect of Premeditation is negative. The effect of Judgmental confidence is positive for females and negative for males; the interaction is significant at p<0.1. The gender dummy is highly insignificant and hence omitted.

[39] Working memory explains 5.5% of the variance in Math, while the personality variables explain additional 12 percentage points. Short-term memory is highly insignificant. The effect of Sensation-seeking is negative. The effects of Risk aversion and Math anxiety are qualitatively the same as for MATHresid1. The gender dummy is highly insignificant and hence omitted. As in the MATHresid1 estimation, Age has a negative effect but is not included in the final estimation.

between Math and Working memory. We reconcile this issue in the multivariate analysis.

Before doing so, we briefly examine determinants of early forecasting performance and learning. Table 1 shows that M_{EARLY} is not correlated with any cognitive variables, although the correlation with Working memory in T_{seq} is almost significant. M_{EARLY} is better for subjects with lower Risk aversion in T_{seq}, and for males in both treatments. We return to the determinants of early performance in the multivariate analysis. Lastly, the extent of learning, M_{EARLY}-M_{LATE}, is positively related to Working memory in both treatments but stronger in T_{seq}. In T_{sim}, learning is greater for subjects with higher Need for cognition and for females.[40]

4.3 Multivariate analysis of forecasting performance

The causality hypothesis proposes that, controlling for the impact of other cognitive, personality and demographic variables, working memory should be a stronger determinant of performance in T_{seq} compared to T_{sim}. To assess the hypothesis, we test for the across-treatment differential in the impact of working memory on M_{LATE} in the following model:

$$^{seq}M_{LATE} - {}^{sim}M_{LATE} = \alpha + X^{seq}\beta^{seq} - X^{sim}\beta^{sim} + (\varepsilon^{seq} - \varepsilon^{sim})$$

All variables are differenced across treatments within subject pairs facing identical Ω_t forecasting sequences.[41] $^{seq}M_{LATE}$ and $^{sim}M_{LATE}$ are the Nx1 performance vectors (N=62), X^{seq} and X^{sim} are the NxK matrices of K individual characteristics included in the estimated model, β^{seq} and β^{sim} are the Kx1 parameter vectors, ε^{seq} and ε^{sim} are the

[40] We caution, however, that the learning correlations are informative only to the extent that M_{EARLY}-M_{LATE} is deemed suitable for comparing the magnitude of learning across subjects with varying values of M_{EARLY}. We alternatively examined various proportional learning metrics and learning speed or duration metrics, but none of them seemed related to the individual characteristics in an economically meaningful way.

[41] The model thus assumes that the effect of Ω_t-complexity on M_{LATE} does not interact with the effect of the included individual characteristics. The assumption seems innocuous given our background results. Namely, we parameterize a broad set of Ω_t-complexity characteristics (variants of those listed in the Appendix) and find in a panel estimation that several of these characteristics weakly influence forecasting performance in early stages of the forecasting task but not in the LATE periods.

regression disturbances, and α is the intercept indicating any remaining across-treatment performance differential.

We estimate the model using a censored Tobit estimator (censored normal regression) that permits "top-bounded" performance to arise in either T_{sim} or T_{seq} within each subject pair.[42] Five subjects in T_{sim} and three in T_{seq} reach M_{LATE}=0, and most of them already have their performance (almost) perfectly top-bounded for quite a while before reaching the LATE periods. The slight across-treatment difference in the extent of top-bounded performance is unlikely to drive the across-treatment differential in the impact of working memory (and other variables) since the presented results hold regardless of including the top-bounded subjects (i.e., pairs). Furthermore, we show for selected models that OLS estimates, while potentially biased due to the censoring issue, yield very similar results.[43]

All individual characteristics are standardized to have zero mean and unit variance, so each coefficient estimate is a (partial) average marginal effect on M_{LATE} of a one-standard-deviation increase in an individual characteristic. In presenting the models, we gradually expand the set of characteristics that are assumed relevant. Due to the different cognitive and possibly also personality (motivational) requirements of T_{sim} and T_{seq}, we *a priori* permit that not only working memory but also other characteristics differ in their impact across treatments. To gain efficiency though, estimates are pooled whenever they do not differ across treatments at $p<0.1$ using a two-tail Wald test. As an exception, we use a one-tail Wald test when testing for the across-treatment differential in the impact of working memory and short-term memory since we hypothesize that their impact is greater in T_{seq}.

Table 2 displays the first set of models where M_{LATE} is regressed on the cognitive abilities themselves rather than their residuals (see the previous section). *Model 1* presents the most bare-bone test of the causality hypothesis. Confirming the correlation

[42] In one case, both subjects in a given pair are top-bounded. We treat this as a no-censoring case with no consequences for the reported results.
[43] The OLS estimation might also be viewed as a useful robustness check in that the censored Tobit estimator is asymptotic and relies on the assumption of i.i.d. normal disturbances which is seemingly met but difficult to test reliably given the small sample size.

results, Working memory significantly fosters performance only in T_{seq} where a one-standard-deviation increase in Working memory corresponds to a performance increase of 52%, compared to an increase of 18% in T_{sim}. The across-treatment differential is significant at $p<0.1$. The insignificant intercept indicates better performance in T_{sim} compared to T_{seq}, in magnitude matching the insignificant across-treatment differential in average M_{LATE} reported in Section 4.1. The intercept generally gets even smaller in the remaining models (including those in Tables 3 and 4), confirming that the cognitive load treatment manipulation has only a weak average effect on asymptotic performance.

In *Model 2*, both higher Working memory and Short-term memory contribute significantly and similarly in magnitude to improving performance in T_{seq}. The size of the Working memory effect is slightly reduced compared to *Model 1*, likely due to the shared variance with Short-term memory. Both Working memory and Short-term memory exhibit a significant across-treatment differential at $p<0.1$, suggesting that they both have a causal effect on performance.

In *Model 3*, which additionally includes Math, the effects of Working memory and Short-term memory remain virtually intact in T_{seq}, and the across-treatment differential for Working memory becomes significant at $p<0.05$. Math turns out to significantly foster performance only in T_{sim} where, controlling for the (now even weaker) impact of Working and Short-term memory, a one-standard-deviation increase in Math is associated with a 43% increase in performance. The Math across-treatment differential just misses significance ($p=0.126$). Hence relaxing the memory load in T_{sim} compared to T_{seq} seems to transform the forecasting task from a memory-intensive into a number-intensive one. The finding that Math affects performance in T_{sim} rather than T_{seq} contradicts the correlation results; the discrepancy most likely arises from the shared variance between Math and Working memory, which the multivariate analysis takes into account.

Controlling for the effect of cognitive abilities, we next examine the effect of personality and demographic characteristics, eventually including only those significantly predicting performance.[44] Overall, their effects are much smaller in magnitude compared to those of Working memory and Short-term memory in T_{seq} and Math in T_{sim}. We first consider Need for cognition since it seems the most theoretically relevant personality determinant of performance. In *Model 4*, higher Need for cognition indeed fosters performance. The Math across-treatment differential becomes significant at $p<0.1$ and stays significant in the remaining models. In *Model* 5, subjects with lower Math anxiety perform better, whereas the impact of Need for cognition becomes smaller and insignificant (we nevertheless keep the variable for model comparison since excluding it makes no difference for the other estimates). The across-treatment differentials for both Working memory and Short-term memory are now significant at $p<0.05$.

Finally, *Model 6* suggests that receiving the windfall bonus further stimulates performance in T_{seq}. The insignificant Windfall effect in T_{sim} has the opposite sign, and the across-treatment differential is significant at $p<0.1$. Despite the bonus being awarded entirely exogenous to the forecasting task, we find in an ordered logit estimation (or alternatively a linear probability model estimation) that the bonus in fact went more often to males, subjects with higher Math score and Risk aversion, lower Perseverance and Math anxiety, and females with higher Sensation-seeking. We check that parsing out these effects from the Windfall score and including only the Windfall residuals, WINDresid, leaves the results of *Model 6* intact. Also, neither Windfall nor WINDresid seem to interact with any cognitive, personality and demographic characteristics.

Model 7 represents an OLS counterpart of *Model 6*. The OLS estimates for Working memory and Short-term memory in T_{seq} and Math in T_{sim} are slightly biased towards zero compared to the censored Tobit estimates in *Model 6*. However, the qualitative conclusions drawn from the two alternative estimations are essentially identical. In *Model* 7, the included regressors together explain 43% of between-subject (and across-treatment) variance in forecasting performance.

[44] The remaining personality and demographic variables as well as any interactions and higher-order moments are individually and jointly insignificant at conventional significance levels (usually highly insignificant), and including them tends to considerably reduce the precision of the reported estimates.

Taken together, the richer *Models 4-7* confirm and in some cases strengthen (in terms of statistical significance) the explanatory power of Working memory and Short-term memory in T_{seq} and Math in T_{sim}. The effect sizes remain remarkably stable across the models, regardless of which other variables are included. In the richest *Model 6*, one-standard-deviation increases in Working memory and Short-term memory in T_{seq} and Math in T_{sim} correspond, *ceteris paribus*, to performance increases of 39%, 40% and 48%, respectively. The across-treatment differentials for Working memory and Short-term memory always reach $p<0.1$ and sometimes $p<0.05$, so the causality hypothesis receives robust and moderately strong support. In the models presented next, the support is even stronger.

In particular, Table 3 displays a second set of models where M_{LATE} is regressed on the residual variation in cognitive abilities. There are three pairs of models, where the left model (e.g., *Model 4a*) features WMresid1 and MATHresid1 from which personality traits were parsed out, while the right model (e.g., *Model 4b*) features WMresid2, STMresid2 and MATHresid2 from which their shared variance and personality traits were parsed out (see the previous section). The first, second and third pairs of models in Table 3 corresponds to *Models 4, 5 and 6* in Table 2, respectively, and were selected using the same criteria.

Comparing the corresponding models across the two tables, the effects of working and short-term memory are qualitatively very similar. Quantitatively, however, the effect of working memory in T_{seq} now becomes substantially larger in magnitude, especially in the models with WMresid2. In the richest *Models 6a* and *6b,* a one-standard-deviation increase in WMresid1 and WMresid2 in T_{seq} corresponds to a performance increase of 43% and 69%, respectively. The effect of STMresid2 in T_{seq} is also larger in magnitude and more significant. In some models, WMresid2 and especially STMresid2 become significant even in T_{sim}, though the effect size is still much smaller compared to T_{seq}. The across-treatment differential for working memory and hence the support for the causality hypothesis becomes stronger, always reaching $p<0.05$ for WMresid1 and $p<0.01$ for WMresid2. The across-treatment differential for STMresid2 is also stronger, always reaching $p<0.05$.

Compared to the effect of Math, the effect of MATHresid1 in T_{sim} is slightly smaller in magnitude but still highly significant; the across-treatment differential stays about the same, always reaching $p<0.1$. By contrast, the across-treatment differential for MATHresid2 is highly insignificant; the pooled effect is smaller in magnitude though still weakly significant. Need for cognition is a strong positive predictor of performance only in *Model 4a*. In the remaining models, we include the variable whenever it just misses $p<0.1$, but we omit it from the richer models with WMresid2 where it becomes highly insignificant and hinders efficiency. Math anxiety now has a slightly larger negative effect on performance, being always significant at $p<0.05$ or better. The windfall bonus now has a weaker positive effect on performance in T_{seq} where it just misses $p<0.1$. We nevertheless include the variable for comparison with the richest models in Table 2, but here we use WINDresid which predicts performance slightly better than Windfall itself. Lastly, the results of the richest censored Tobit *Model 6b* are confirmed in the counterpart OLS *Model* 7b, which further shows that the included regressors together explain 40% of between-subject (and across-treatment) variance in forecasting performance.

The last set of models in Table 4 attempts to additionally control for the influence of prior forecasting expertise. We have shown that especially in T_{sim}, M_{EARLY} and M_{LATE} correlate noticeably stronger with each other than either of them separately correlates with individual characteristics. Hence both M_{EARLY} and M_{LATE} might be influenced by unobserved "intrinsic forecasting ability," and not taking this into account might bias the conclusions about the impact of the measured characteristics on performance. As a precaution against such a possibility, we create a proxy for intrinsic forecasting ability and include it in the M_{LATE} estimation. An efficient way to create the proxy is to extract it from M_{EARLY} performance which is based on the same segment of Ω_t as is M_{LATE}. Our proxy is therefore the residual variance in M_{EARLY} that remains after parsing out theoretically and statistically relevant individual characteristics. In that way, the effect of the proxy on M_{LATE} will not reflect the impact of those characteristics, so they should retain their independent influence on M_{LATE} if any.

We in fact create two proxies. EARLYresid1 are the residuals extracted from a OLS regression of M_{EARLY} on cognitive abilities (of which only Math in T_{sim} significantly fosters M_{EARLY}) as well as statistically significant personality and demographic characteristics (namely, M_{EARLY} is better at $p<0.1$ for males and for subjects with lower Sensation-seeking in T_{sim}). On the other hand, EARLYresid2 are the residuals from a OLS regression on cognitive abilities only, of which Working memory in T_{seq} and Math in T_{sim} significantly improve performance.[45] In Table 4, the models are again arranged in pairs, where the left model (e.g., *Model 4c*) includes EARLYresid1 while the right model (e.g., *Model 4d*) includes EARLYresid2. All models feature WMresid2, STMresid2 and MATHresid2 (see the previous section). Hence the first, second and third pairs of models in Table 4 corresponds to *Models 4b, 5b and 6b* in Table 3, respectively, and were selected using the same criteria.

Comparing the corresponding models across the two tables, the effects of working memory and short-term memory are both qualitatively and quantitatively very similar. Also, there are only minor quantitative differences between the paired models in Table 4. The across-treatment differential for WMresid2 is with one exception significant at $p<0.01$, while the differential for STMresid2 always reaches $p<0.05$. The pooled effect of MATHresid2 becomes more significant but stays about the same in magnitude. Need for cognition again fosters performance only in the simpler *Models 4c* and *4d* but it becomes highly insignificant in the richer models from which it is omitted. The negative effect of Math anxiety on performance stays the same and highly significant. As in *Model 6b*, windfall bonus has a weak positive effect on performance in T_{seq} where it just misses $p<0.1$. We again include the variable for model comparison and use WINDresid which predicts performance slightly better than Windfall itself.

[45] M_{EARLY} performance is not top-bounded for any subject, permitting the use of OLS. In order to retain the richest possible model, we use the measured cognitive abilities rather than their residuals, and we do not allow parameters to be pooled across treatments. The M_{EARLY} estimation is conducted in the paired manner as for M_{LATE}. Thus including EARLYredis1 or EARLYresid2 in the M_{LATE} estimation assumes that their impact does not differ across treatments. The assumption seems empirically warranted but is hard to test due to the considerable loss of efficiency when the M_{EARLY} estimation is alternatively conducted in an unpaired manner. The negative impact of Sensation-seeking on M_{EARLY} is reflected in *Models 5c* and *6c* where the variable has a positive significant impact on M_{LATE} in T_{sim}. This "compensation" effect is of secondary importance and thus is not reported in Table 4.

The main novel insight from the models in Table 4 is that intrinsic forecasting ability is a strong positive predictor of performance. The effect of EARLYresid1 is always slightly smaller in magnitude compared to the effect of EARLYresid2, most likely due to parsing out the personality and demographic characteristics from the former. The OLS *Model 7d*, besides confirming the results of the counterpart censored Tobit *Model 6d*, shows that including EARLYresid2 in the estimation raises the explained share of between-subject (and across-treatment) variance in performance to 57% (compared to 40% in *Model 7b*).

5. Conclusion

Using a memory-intensive time-series forecasting task, we identify the causal effect of working memory and short-term memory on forecasting performance. The higher memory load in the sequential presentation treatment "activates" subjects' working and short-term memory constraints. The constraints become irrelevant in the less memory-intensive simultaneous presentation treatment where the explanatory power shifts to basic arithmetic abilities. By removing the shared variance between working and short-term memory, we show that their causal effects are independent of each other, and in turn that the causality of working memory likely reflects individual heterogeneity in the capacity to control attention. Since this capacity is a strong and reliable predictor of behavior in a wide range of psychology tasks requiring controlled information processing, our results suggest that the capacity might also influence decision quality in cognitively complex economic settings. Naturally, the predictive power of the capacity might vary, for instance, with task complexity, educational level, and over the life cycle, as seems to be the case for broader measures of general intelligence (Almlund et al., 2011).

Controlling for the effect of theoretically relevant personality traits, we find that forecasting performance is better for subjects with lower math anxiety and, to a weaker extent, higher intrinsic motivation. Further exploring the role of motivational factors, a large windfall financial bonus won prior to the forecasting task fosters performance in the sequential presentation treatment. Nevertheless, merely controlling for the impact of personality and motivational factors constitutes only an initial step in examining their interaction with cognitive abilities, and, more generally, in studying the multitude of structural relationships that cognitive production likely entails. Below we discuss some of the relationships and how we plan to address them in the forecasting setting, with the prerequisite of having established the causality of working and short-term memory.

To start with, the channels behind the causal effect of cognitive abilities might be numerous. In our setting, for example, working memory might influence not only subjects' ability to effectively combine forecast-relevant information but also the nature of their forecasting strategies (e.g., Barrick and Spilker, 2003; Libby and Luft, 1993; strategies might also be affected by the memory load manipulation itself). In the Appendix, we make casual observations regarding the importance of the strategy channel. Psychologists have further argued that not only the objective cognitive abilities but also their self-perception and confidence in them (self-efficacy) may influence behavior (e.g., Bandura and Locke, 2003). One way of exploring the confidence channel in our forecasting setting (using a larger data set) is to interpret subjects' bets as a measure of confidence in forecasting abilities. After removing the effect of personality traits such as risk aversion and *ex ante* judgmental confidence from the bets, one can examine whether this measure of confidence fosters performance beyond the direct effect of forecasting abilities themselves, and how the confidence depends on subjects' cognitive and personality characteristics.

One of economically relevant relationships in cognitive production is the degree of substitutability among cognitive abilities varying in task specificity. We examined the predictive power of three cognitive abilities varying from domain-general (working memory) to more task-specific (short-term memory and especially basic arithmetic abilities). We also saw their predictive power to persist even after subjects extensively acquired task-specific cognitive ability in the form of on-task experience. This is in line with evidence suggesting that various forms of on-task experience do not harm the predictive power of working memory (Engle and Kane, 2004). Nevertheless, we did not explicitly consider the role of task-specific cognitive ability in the form of prior forecasting expertise (or domain knowledge), although we did create a proxy for it based on early forecasting performance. There exists only preliminary evidence regarding the interplay between prior expertise and general cognitive abilities in economic settings (Hambrick and Engle, 2003). As an initial step, Wittmann and Süß (1999) document that both domain knowledge and working memory correlate positively with performance in a simulated physical production task. Closer to our setting, Ghosh and Whitecotton (1997) find that higher perceptual abilities are associated with better performance in a company earnings prediction task, and that this effect is overcome neither by prior expertise of professional financial analysts nor by provision of a decision aid.[46]

Arguably, however, only after establishing the causal effect of general cognitive abilities can one credibly assess their substitutability with prior expertise. Our forecasting task lends itself to examining that substitutability since it naturally extends to a real-world setting. Namely, one could adapt the presently abstract task into a financially framed version, for instance by interpreting the forecasted variable Ω_t as a commodity price following our (or similar) seasonal process, and B_t as an economically relevant, perfectly predictable state variable linearly related to Ω_t. Using the sequential and simultaneous presentation treatments, one could then challenge inexperienced forecasters (students) and experienced forecasters (e.g., commodity traders) with the framed and unframed versions of the task. The resulting 2x2x2 between-subjects design would shed further light on the above established causality of working and short-term memory and permit direct assessment of their substitutability with prior expertise.

Leaving the confines of cognitive abilities and getting to the heart of cognitive production, one naturally turns to the interplay between cognitive ability and cognitive effort. Evidence from the working memory literature is suggestive of a limited degree of ability-effort substitutability. In tasks where working memory is a strong predictor of performance, effort latencies (inferred from response times, pupil dilation, fMRI data, etc.) do not vary across people with different working memory, while effort latencies tend to increase relatively uniformly with higher financial incentives and task

[46] The company earnings prediction literature generally provides inconclusive evidence regarding the role of prior expertise, both in the lab and the field (e.g., Hunton and McEwen, 1997; Libby et al., 2002).

complexity (e.g., Heitz et al., 2008). Awasthi and Pratt (1990) provide further circumstantial evidence of limited ability-effort substitutability, in that piece-rate (as compared to flat-wage) financial incentives yield better judgmental performance only for individuals with higher perceptual abilities while effort duration increases uniformly regardless of the abilities. These observations suggest that, even if financial incentives and personality (motivational) traits sufficiently induce effort, both financial and cognitive resources might be wasted for people lacking cognitive abilities to perform cognitively demanding tasks. This basic prediction, expressed in Camerer and Hogarth's (1999) capital-labor-production framework and in various modifications by others,[47] seems to have remained empirically unexplored.[48]

A potential explanation is that cognitively more constrained individuals are less aware of their "objective" cognitive abilities (e.g., Critcher and Dunning, 2009; Ehrlinger et al., 2008). Indeed, the working memory literature suggests that people with lower working memory relying predominantly on automated processing possess noisier estimates of their abilities, compared to people with higher working memory relying mostly on controlled processing (e.g., Feldman-Barrett et al., 2004). One way of explicitly exploring this issue is to examine whether and which people are willing to pay for the relaxation of their cognitive constraints. In the forecasting setting, imagine that subjects start forecasting in the more memory-intensive sequential presentation treatment, but can pay (at the start of each period) for switching to the less memory-intensive simultaneous presentation treatment. Subjects could therefore choose to purchase "external" memory. Figure 1 illustrates that switching to the simultaneous presentation treatment does not guarantee perfect performance but it does improve performance and learning progress on average. Subjects of course do not know this and their switching decisions would presumably reflect their expectation that the net (long-run) return to switching is positive. Thus similar to bets, switching behavior yields a

[47] See, e.g., Almlund et al. (2011), Awasthi and Pratt (1990), Conlisk (1980), Libby and Lipe (1992), Libby and Luft (1993), Segal (2008), and Wilcox (1993).

[48] Indirect evidence from meta-studies and empirical surveys indicates that incentive effects depend in a complicated fashion on the nature of cognitive tasks (e.g., Bonner et al., 2000; Camerer and Hogarth, 1999; Hertwig and Ortmann, 2001, 2003; Jenkins et al., 1998). Promising initial steps towards understanding incentive effects have also been made in neurobiology (e.g., Gold and Shadlen, 2001) and neuroeconomics (e.g., Camerer et al., 2005).

decision-relevant and incentive-compatible indicator of subjects' estimates of their forecasting abilities, which can in turn be linked to the measured cognitive and personality characteristics and forecasting performance. One may further want to examine the effect of varying the cost of switching.

A different perspective on ability-effort substitutability may be gained by focusing on the (reduced-form) interaction among cognitive abilities, personality traits and financial incentives. As mentioned in Section 1, establishing the causality of cognitive abilities is a prerequisite for addressing issues relevant for the design of efficient incentive schemes, such as how people behave under different incentive levels and schemes conditional on their abilities (and traits), and how they self-select into the schemes based on their abilities (and traits). Initial experimental evidence suggests a positive interaction of incentives and abilities in fostering performance. In an accounting task, Awasthi and Pratt (1990) find that incentives (rewarding correct choice) improve judgmental performance compared to flat-wage pay only for subjects with high perceptual abilities. Similarly, Palacios-Huerta (2003) reports that increasing incentives (piece-rate and tournament) improves performance in a Monty Hall Three Door task only for subjects with superior schooling outcomes. Nevertheless, this evidence is purely correlational and does not enable one to condition on cognitive abilities. By contrast, we can credibly study incentive variation in the forecasting task while conditioning on working memory and short-term memory. We could also interact incentive variation with the cognitive load variation, interpreting the simultaneous and sequential presentation treatments as less and more demanding work settings, respectively.

One could further explore the welfare (performance) implications of implementing the forecasting task under various incentive schemes – say, the presently used piece-rate scheme, a quota scheme, a tournament scheme and a flat-wage scheme. Due to their varying returns to cognitive abilities and effort as well as varying degree of competitiveness, the incentive schemes likely differ in the extent to which abilities and traits predict performance. One could examine whether financial incentives can be utilized efficiently by *ex ante* assigning employees to incentives schemes which best correspond to their observed (measured) abilities and traits. A natural extension is to examine how people self-select based on their abilities and traits into the incentive schemes and whether this endogenous sorting is efficient compared to the exogenous assignment. Recent experimental literature has shown that endogenous sorting can be efficient (in terms of output) and seems driven mainly by productivity sorting and partly by personality traits such as risk attitudes (e.g., Dohmen and Falk, 2011; Cadsby et al., 2007). Again, however, the literature has not attended systematically to cognitive and personality determinants of performance and sorting, and especially to the issue of causality.

To conclude, economists widely believe that, absent strategic considerations such as agency problems, financial incentives represent a dominant and effective stimulator of human productive activities. In production settings that are cognitively demanding, however, the effectiveness of financial incentives may be moderated by individual cognitive abilities and personality traits. We provide initial evidence from an economic setting that the effectiveness of even strong financial incentives may be moderated by cognitive abilities in a causal fashion. In line with the correlational evidence reviewed in Section 1, our findings illustrate the need to attend to cognitive constraints, besides personality and preference-based factors, when interpreting observed (variance of) behavior in cognitively demanding lab and field economic environments.

Role of the Funding Source

The research was conducted with the financial support from the Grant Agency of the Czech Republic (grant No. 402/05/1023). The grant provided funds for developing the study design, data collection (payments or experimental subjects), initial analysis and interpretation of data, and conference travel. Further travel funds were obtained from the Hlavka Foundation, Prague, and CERGE-EI, Prague.

References

Ackerman, P.L., Beier, M.E., Boyle, M.O., 2002. Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. Journal of Experimental Psychology: General 131, 567-589.

Almlund, M., Duckworth, A.L., Heckman, J.J., Kautz, T., 2011. Personality psychology and economics. In: Hanushek, E.A., Machin, S.J., Woessmann, L. (Eds.). Handbook of the Economics of Education, Vol. IV. Amsterdam: Elsevier, 1-181.

Anderson, J. R., 2000. Cognitive Psychology and its Implications. New York: Worth.

Awasthi, V., Pratt, J., 1990. The effects of monetary incentives on effort and decision performance: The role of cognitive characteristics. Accounting Review 65, 797-811.

Ballinger, P.T., Hudson, E., Karkoviata, L., Wilcox, N.T., 2011. Saving behavior and cognitive abilities. Experimental Economics 14, 349-374.

Bandura, A., Locke, E.E., 2003. Negative self-efficacy and goal effects revisited. Journal of Applied Psychology 88, 87-99.

Barrick, J.A., Spilker, B.C., 2003. The relations between knowledge, search strategy, and performance in unaided and aided information search. Organizational Behavior and Human Decision Processes 90, 1-18.

Benjamin, D.J., Brown, S.A., Shapiro, J.M., Unpublished results. Who is ‘behavioral’? Cognitive ability and anomalous preferences. Harvard University manuscript.

Ben-Ner, A., Kramer, A., Levy, O., 2008. Economic and hypothetical dictator game experiments: Incentive effects at the individual level. Journal of Socio-Economics 37, 1775-1784.

Bonner, S.E., Hastie, R., Sprinkle, G.B., Young, S.M., 2000. A Review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. Journal of Management Accounting Research 12, 19-64.

Bonner, S.E., Sprinkle, G.B., 2002. The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. Accounting, Organizations & Society 27, 303-345.

Borghans, L., Meijers, H., Ter Weel, B., 2008. The role of noncognitive skills in explaining cognitive test scores. Economic Inquiry 46, 2-12.

Brehmer, B., 1980. In one word: Not from experience. Acta Psychologica 45, 223-241.

Burnham, T.C., Cesarini, D., Johannesson, M., Lichtenstein, P., Wallace, B., 2009. Higher cognitive ability is associated with lower entries in a p-beauty contest. Journal of Economic Behavior and Organization 72, 171-175.

Cacioppo, J.T., Petty, R.E., Kao, C.F., 1984. The efficient assessment of need for cognition. Journal of Personality Assessment 48, 306-307.

Cacioppo, J.T., Petty, R.E., Feinstein, J.A., Jarvis, W.B.G., 1996. Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. Psychological Bulletin 119, 197-253.

Cadsby, C.B., Song, F., Tapon, F., 2007. Sorting and incentive effects of pay for performance: An experimental investigation. Academy of Management Journal 50, 387-405.

Camerer, C.F., Hogarth, R., 1999. The effects of financial incentives in experiments: A review and capital-labor-production framework. Journal of Risk and Uncertainty 19, 7-42.

Camerer, C.F., Loewenstein, G., Prelec, D., 2005. Neuroeconomics: How neuroscience can inform economics. Journal of Economic Literature 34, 9-64.

Cawley, J., Heckman, J.J., Vytlacil, E.J., 2001. Three observations on wages and measured cognitive ability. Labour Economics 8, 419-442.

Conlisk, J., 1980. Costly optimization versus cheap imitators. Journal of Economic Behavior and Organization 1, 275-293.

Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., Engle, R.W., 2005. Working memory span tasks: A methodological review and user's guide. Psychonomic Bulletin and Review 12, 769-786.

Cooper, S., Robinson, D., 1991. The relationship of mathematics self-efficacy beliefs to mathematics anxiety and performance. Measurement and Evaluation in Counseling and Development 24, 4-11.

Critcher, C.R., Dunning, D., 2009. How chronic self-views influence (and mislead) self-evaluations of performance: Self-views shape bottom-up experiences with the task. Journal of Personality and Social Psychology 97, 931-945.

Crowley, A.E., Hoyer, W.D., 1989. The relationship between need for cognition and other individual difference variables: A two-dimensional framework. Advances in Consumer Research 16, 37-43.

Cunha, F., Heckman, J.J., Schennach, S.M., 2010. Estimating the technology of cognitive and noncognitive skill formation. Econometrica 78, 883-931.

Dave, Ch., Eckel, C.C., Johnson, C.A., Rojas, Ch., 2010. Eliciting risk preferences: When is simple better? Journal of Risk and Uncertainty 41, 219-243.

Dawes, R.M., 1979. The robust beauty of improper linear models in decision making. American Psychologist 34, 371-582.

Deci, E., Koestner, R., Ryan, R., 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological Bulletin 125, 627-668.

Devetag, G., Warglien, M., 2003. Games and phone numbers: Do short term memory bounds affect strategic behavior? Journal of Economic Psychology 24, 189-202.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., 2010. Are risk aversion and impatience related to cognitive ability? American Economic Review 100, 1238-1260.

Dohmen, T., Falk, A., 2011. Performance pay and multidimensional sorting: Productivity, preferences, and gender. American Economic Review 101, 556-590.

Duffy, S., Smith, J., Unpublished results. Cognitive load in the multi-player prisoner's dilemma game: Are there brains in games? Rutgers University-Camden manuscript.

Dwyer, G., Williams, A., Battalio, R., Mason, T., 1993. Tests of rational expectations in a stark setting. The Economic Journal 103, 586-601.

Eckel, C., Wilson, R., 2004. Is trust a risky decision? Journal of Economic Behavior and Organization 55, 447-465.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., Kruger, J., 2008. Why the unskilled are unaware? Further explorations of (lack of) self-insight among the incompetent. Organizational Behavior and Human Decision Processes 105, 98-121.

Eisenberger, R., Cameron, J., 1996. Detrimental effects of reward: Reality or myth? American Psychologist 51, 1153-1166.

Ekstrom, R.B., French, J.W., Harman, H., Derman, D., 1976. Kit of Factor-Referenced Cognitive Tests (rev. ed.). Princeton, NJ: Educational Testing Service.

Engle, R.W., Tuholski, S.W., Laughlin, J.E., Conway, A.R.A., 1999. Working memory, short-term memory, and general fluid intelligence: A latent variable approach. Journal of Experimental Psychology: General 128, 309-331.

Engle, R.W., Kane, M.J., 2004. Executive attention, working memory capacity, and a two-factor theory of cognitive control. In: Ross, B. (Ed.). The Psychology of Learning and Motivation, Vol. XLIV. NY: Elsevier, 145-199.

Ericsson, K.A., Smith, J. (Eds.), 1991. Toward a General Theory of Expertise: Prospects and Limits. Cambridge, UK: Cambridge University Press.

Feldman-Barrett, L., Tugade, M.M., Engle, R.W., 2004. Individual differences in working memory capacity and dual-process theories of the mind. Psychological Bulletin 130, 553-573.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10, 171-178.

Frederick, S., 2005. Cognitive reflection and decision making. Journal of Economic Perspectives 19, 25-42.

Ghosh, D., Whitecotton, S.M., 1997. Some determinants of analysts' forecast accuracy. Behavioral Research in Accounting 9 (Supplement), 50-68.

Gneezy, U., Rustichini, A., 2000. Pay enough or don't pay at all. Quarterly Journal of Economics 115, 791-811.

Gold, J.I., Shadlen, M.N., 2001. Neural computations that underlie decisions about sensory stimuli. Trends in Cognitive Sciences 5, 10-16.

Hambrick, D.Z., Engle, R.W., 2003. The role of working memory in problem solving. In: Davidson, J.E., Sternberg, R.J. (Eds.). The psychology of Problem Solving. London: Cambridge Press, 176-206.

Hammond, K.R., McClelland, G.H., Mumpower, J., 1980. Human Judgment and Decision Making: Theories, Methods, and Procedures. New York: Praeger.

Harrison, G.W., List, J.A., 2004. Field experiments. Journal of Economic Literature 42, 1009-1055.

Harrison, G.W., Rutström, E.E., 2008. Experimental evidence on the existence of hypothetical bias in value elicitation methods. In: Plott, C.R., Smith V.L. (Eds.).

Handbook of Experimental Economics Results, Vol. I. Amsterdam: North-Holland, 752-767.

Harvey, N., Bolger, F., McClelland, A., 1994. On the nature of expectations. British Journal of Psychology 85, 203-229.

Heckman, J.J., Stixrud, J., Urzua, S., 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. Journal of Labor Economics 24, 411-482.

Heckman, J.J., Malofeeva, L., Pinto, R., Savelyev, P.A., Unpublished results. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. University of Chicago manuscript.

Heitz, R.P., Unsworth, N., Engle, R.W., 2005. Working memory capacity, attentional control, and fluid intelligence. In: Wilhelm, O., Engle, R.W. (Eds.). Handbook of Understanding and Measuring Intelligence. London: Sage Publications, 61-78.

Heitz, R.P., Schrock, J.C., Payne, T.W., Engle, R.W., 2008. Effects of incentive on working memory capacity: Behavioral and pupillometric data. Psychophysiology 45, 119-129.

Hertwig, R., Ortmann, A., 2001. Experimental practices in economics: A methodological challenge for psychologists? Behavioral and Brain Sciences 24, 383-451.

Hertwig, R., Ortmann, A., 2003. Economists' and psychologists' experimental practices: How they differ, why they differ, and how they could converge. In: Brocas, I., Carrillo, J.D. (Eds.). The Psychology of Economic Decisions, Vol. I. Oxford: OUP, 253-272.

Hey, J.D., 1994. Expectations formation: Rational or adaptive or ...? Journal of Economic Behavior and Organization 25, 329-349.

Hinson, J.M., Jameson, T.L., Whitney, P., 2003. Impulsive decision making and working memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 29, 298-306.

Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. American Economic Review 92, 1644-55.

Hunton, J.E., McEwen, R.A., 1997. An assessment of the relation between analysts' earnings forecast accuracy, motivational incentives and cognitive information search strategy. The Accounting Review 72, 497-515.

Isen, A.M., 2000. Positive affect and decision making. In: Lewis, M., Haviland-Jones, J. (Eds.). Handbook of Emotions. New York: Guilford, 417-435.

Isen, A.M., Reeve, J., 2006. The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. Motivation and Emotion 29, 295-323.

Jenkins, D., Jr., Mitra, A., Gupta, N., Shaw, J., 1998. Are financial incentives related to performance? A meta-analytic review of empirical research. Journal of Applied Psychology 83, 777-787.

Kagel, J.H., Levin, D., 1986. The winner's curse and public information in common value auctions. American Economic Review 76, 894-920.

Kahneman, D., Tversky, A., 1984. Choices, values, and frames. American Psychologist 39, 341-350.

Kane, M.J., Bleckey, M.K., Conway, A.R.A., Engle, R.W., 2001. A controller-attention view of working-memory capacity. Journal of Experimental Psychology: General 130, 169-183.

Kane, M.J., Engle, R.W., 2003. Working memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. Journal of Experimental Psychology: General 132, 47-70.

Kane, M.J., Hambrick, D.Z., Tuholski, S.W., Wilhelm, O., Payne, T.W., Engle, R.W., 2004. The generality of working memory capacity: A latent variable approach to verbal and visuospatial memory span and reasoning. Journal of Experimental Psychology: General 133, 189-217.

Kellogg, C.E., Morten, N.W., 1999. Beta III Manual. San Antonio, TX: The Psychological Corporation.

Klayman, J., 1984. Learning from feedback in probabilistic environments. Acta Psychologica 56, 81-92.

Klayman, J., 1988. Cue discovery in probabilistic environments: Uncertainty and experimentation. Learning, Memory, and Cognition 14, 317-330.

Lawrence, M., O’Connor, M., 2005. Judgmental forecasting in the presence of loss functions. International Journal of Forecasting 21, 3-14.

Libby, R., Lipe, M.G., 1992. Incentives, effort, and the cognitive processes involved in accounting-related judgments. Journal of Accounting Research 30, 249-273.

Libby, R., Luft, J., 1993. Determinants of judgment performance in accounting settings: Ability, knowledge, motivation, and environment. Accounting, Organizations & Society 18, 425-450.

Libby, R., Bloomfield, R., Nelson, M., 2002. Experimental research in financial accounting. Accounting, Organizations & Society 27, 775-810.

Maines, L., Hand, J., 1996. Individuals’ perceptions and misperceptions of time series properties of quarterly earnings. Accounting Review 71, 317-336.

McDaniel, T,M., Rutström, E.E., 2001. Decision making costs and problem solving performance. Experimental Economics 4, 145-161.

Mirels, H.L., Greblo, P., Dean, J.B., 2002. Judgmental self-doubt: beliefs about one’s judgmental prowess. Personality and Individual Differences 33, 741-758.

Murayama K., Matsumoto, M., Izuma, K., Matsumoto, K., 2010. Neural basis of the undermining effect of monetary reward on intrinsic motivation. Proceedings of the National Academy of Sciences of the United States of America 107, 20911-20916.

Oechssler, J., Roider, A., Schmitz, P.W., 2009. Cognitive abilities and behavioural biases. Journal of Economic Behavior and Organization 72, 147-152.

Pajares, F., Miller, M.D., 1994. The role of self-efficacy and self-concept beliefs in mathematical problem-solving: A path analysis. Journal of Educational Psychology 86, 193-203.

Pajares, F., Urdan, T., 1996. An exploratory factor analysis of the mathematics anxiety scale. Measurement and Evaluation in Counseling and Development 29, 35-47.

Palacios-Huerta, I., 2003. Learning to open Monty Hall's doors. Experimental Economics 6, 235-251.

Raven, J., Raven, J. C., Court, J.H., 1998. Manual for Raven's Progressive Matrices and Vocabulary Scales. San Antonio, TX: The Psychological Corporation.

Reber, A. S., 1989. Implicit learning and tacit knowledge. Journal of Experimental Psychology: General 118, 219-235.

Rydval. O., Ortmann, A., Ostatnicky, M., 2009. Three very simple games and what it takes to solve them. Journal of Economic Behavior and Organization 72, 589-601.

Schneider, W., Eschman, A., Zuccolotto, A., 2002. E-prime User's Guide. Pittsburgh: Psychology Software Tools.

Schwarzer, R., Seipp, B., Schwarzer, C., 1989. Mathematics performance and anxiety: A meta-analysis. In: Schwarzer, R., Van Der Ploeg, H.M., Spielberger, C.D. (Eds.). Advances in Test Anxiety Research, Vol. VI. Berwyn, PA: Swets North America, 105-119.

Segal, C., Unpublished results. Motivation, test scores, and economic success. Universitat Pompeu Fabra manuscript.

Shiv, B., Fedorikhin, A., 1999. Heart and mind in conflict: The interplay of affect and cognition in consumer decision making. Journal of Consumer Research 26, 278-292.

Stanovich, K.E., West, R.F., 2000. Individual differences in reasoning: Implications for the rationality debate? Behavioral and Brain Sciences 23, 645-665.

Stevens, D., Williams, A., 2004. Inefficiency in earnings forecasts: Experimental evidence of reactions to positive vs. negative information. Experimental Economics 7, 75-92.

Toplak, M.E., West, R.F., Stanovich, K.E., 2011. The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. Memory and Cognition 39, 1275-1289.

Turner, M.L., Engle, R.W., 1989. Is working memory capacity task-dependent? Journal of Memory and Language 28, 127-154.

Welsh, M.C., Satterlee-Cartmell, T., Stine, M., 1999. Towers of Hanoi and London: Contribution of working memory and inhibition to performance. Brain and Cognition 41, 231-242.

Whiteside, S.P., Lynam, D.R., 2001. The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. Personality and Individual Differences 30, 669-689.

Wilcox, N., 1993. Lottery choice: Incentives, complexity, and decision time. The Economic Journal 103, 1397-1417.

Wittmann, W.W., Süß, H.M., 1999. Investigating the paths between working memory, intelligence, knowledge, and complex problem solving performances via Brunswik-symmetry. In: Ackerman, P.L., Kyllonen, P.C., Roberts, R.D. (Eds.). Learning and Individual Differences. Process, Trait, and Content Determinants. Washington, D.C.: APA-Books.

Table 1: Correlations between forecasting performance and cognitive, personality and demographic variables in T_{sim} and T_{seq}

	M_{LATE}		M_{EARLY}		M_{EARLY}-M_{LATE}	
Variable	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}
Working memory	-0.109 (0.400)	-0.479*** (0.000)	0.009 (0.946)	-0.198 (0.124)	0.187 (0.145)	0.229* (0.074)
WMresid1	-0.038 (0.767)	-0.447*** (0.000)	0.049 (0.705)	-0.155 (0.231)	0.133 (0.303)	0.253** (0.048)
WMresid2	-0.031 (0.809)	-0.313** (0.013)	0.111 (0.390)	-0.052 (0.686)	0.137 (0.290)	0.261** (0.041)
Short-term memory	-0.045 (0.730)	-0.251** (0.049)	-0.085 (0.512)	-0.029 (0.826)	-0.117 (0.364)	0.163 (0.205)
STMresid2	-0.092 (0.477)	-0.013 (0.920)	-0.137 (0.287)	0.021 (0.873)	-0.127 (0.324)	-0.013 (0.918)
Math	-0.100 (0.440)	-0.214* (0.096)	-0.187 (0.145)	-0.190 (0.139)	-0.109 (0.400)	-0.038 (0.771)
MATHresid1	-0.088 (0.499)	-0.160 (0.216)	-0.163 (0.207)	-0.115 (0.375)	-0.140 (0.276)	-0.039 (0.761)
MATHresid2	-0.067 (0.604)	-0.015 (0.906)	-0.133 (0.303)	-0.032 (0.807)	-0.128 (0.322)	-0.083 (0.521)
Need for cognition	-0.090 (0.488)	-0.105 (0.417)	0.045 (0.728)	0.028 (0.831)	0.218* (0.089)	0.095 (0.465)
Perseverance	-0.110 (0.396)	0.002 (0.990)	-0.024 (0.855)	0.057 (0.658)	0.095 (0.462)	-0.033 (0.802)
Premeditation	-0.212* (0.098)	-0.019 (0.884)	-0.093 (0.471)	-0.081 (0.533)	0.104 (0.421)	-0.105 (0.419)
Sensation-seeking	0.081 (0.533)	-0.056 (0.665)	0.178 (0.166)	0.085 (0.513)	0.130 (0.313)	0.155 (0.230)
Risk aversion	0.118 (0.363)	0.152 (0.238)	0.016 (0.905)	0.243* (0.057)	-0.195 (0.129)	0.059 (0.651)
Math anxiety	0.223* (0.082)	0.143 (0.267)	0.169 (0.190)	0.181 (0.160)	-0.088 (0.495)	0.005 (0.971)
Judgmental confidence	0.023 (0.860)	0.201 (0.117)	-0.024 (0.853)	-0.020 (0.880)	-0.015 (0.909)	-0.133 (0.304)
Age	-0.114 (0.377)	-0.115 (0.372)	-0.131 (0.312)	-0.164 (0.202)	0.009 (0.945)	-0.042 (0.743)
Male	-0.112 (0.387)	-0.283** (0.026)	-0.367*** (0.003)	-0.265** (0.037)	-0.333*** (0.008)	-0.016 (0.900)
Windfall	0.047 (0.715)	-0.160 (0.213)	0.011 (0.934)	0.051 (0.692)	-0.093 (0.472)	0.182 (0.158)

Note: The variables are defined in Section 4. The table displays Spearman rank correlations with *p*-values in parentheses. *,**, and *** indicate significance of the estimates at $p<0.1$, $p<0.05$ and $p<0.01$, respectively.

Table 2: Estimation of forecasting performance on cognitive, personality and demographic characteristics

	Model 1		*Model 2*		*Model 3*		*Model 4*		*Model 5*		*Model 6*		*Model 7*	
REGRESSOR	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}
intercept	1.020		0.893		0.537		0.309		0.578		0.620		0.620	
	(1.072)		(1.021)		(1.039)		(1.034)		(1.033)		(1.015)		(1.015)	
Working memory	-1.017	-3.572**	-0.573	-2.909**	-0.287	-2.973**	-0.352	-2.876***	-0.598	-2.807***	-0.688	-2.706**	-0.744	-2.507***
	(0.967)	(1.382)	(0.853)	(1.364)	(0.621)	(1.190)	(0.601)	(1.060)	(0.597)	(1.051)	(0.564)	(1.034)	(0.599)	(0.891)
	*		*		**		**		**		*		*	
Short-term memory			-0.774	-3.039**	-0.397	-2.908**	-0.463	-2.665**	-0.478	-2.796**	-0.736	-2.757**	-0.851	-2.618**
			(0.661)	(1.216)	(0.647)	(1.209)	(0.620)	(1.170)	(0.637)	(1.083)	(0.579)	(1.096)	(0.554)	(1.146)
			*		*		*		**		*		*	
Math					-2.498**	-0.443	-2.755***	-0.293	-2.510**	0.102	-2.774***	0.435	-2.676***	0.483
					(1.036)	(1.306)	(0.992)	(1.275)	(0.964)	(1.335)	(0.941)	(1.325)	(0.934)	(1.304)
							*		*		**		**	
Need for cognition							-1.186*		-0.693		-0.733		-0.781	
							(0.673)		(0.657)		(0.578)		(0.593)	
Math anxiety									1.437*		1.520*		1.534*	
									(0.811)		(0.782)		(0.833)	
Windfall											0.630	-1.336**	0.814	-1.335*
											(0.915)	(0.638)	(0.837)	(0.682)
											*		**	
Log likelihood	-203.861		-201.365		-198.205		-196.861		-195.371		-193.928		R^2=0.432	

Note: Forecasting performance is M_{LATE}. The regressors are defined in Section 4. Observations = 124, 62 in T_{sim} and 62 in T_{seq}. The displayed estimates are average marginal effects from censored normal regressions, with heteroskedasticity-robust standard errors in parentheses; the last model presents OLS estimates. *,**, and *** indicate significance of estimates as well as their across-treatment differentials (underneath the standard errors) at p<0.1, p<0.05 and p<0.01, respectively, using a two-tail Wald test. As an exception, the across-treatment differentials for Working memory and Short-term memory are compared using a one-tail Wald test. The included regressors are always jointly highly significant.

Table 3: Estimation of forecasting performance on cognitive "residuals" and personality and demographic characteristics

	Model 4a		*Model 4b*		*Model 5a*		*Model 5b*		*Model 6a*		*Model 6b*		*Model 7b*	
REGRESSOR	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}
intercept	0.409		0.618		0.693		0.986		0.680		0.971		0.815	
	(1.029)		(1.057)		(0.999)		(0.987)		(0.994)		(0.980)		(0.950)	
WMresid1 or	-0.330	-3.114***	-0.866	-4.624***	-0.533	-3.148***	-1.019	-4.890***	-0.774	-2.987***	-1.351*	-4.799***	-1.385*	-4.432***
WMresid2	(0.603)	(1.047)	(0.770)	(1.068)	(0.583)	(1.054)	(0.702)	(1.141)	(0.590)	(1.048)	(0.717)	(1.130)	(0.710)	(0.994)
	**		***		**		***		**		***		**	
STMresid1 or	-0.557	-2.665**	-1.331*	-4.164***	-0.594	-2.791**	-1.393**	-4.537***	-0.697	-2.888***	-1.681***	-4.659***	-1.740***	-4.344***
STMresid2	(0.584)	(1.178)	(0.675)	(1.328)	(0.586)	(1.063)	(0.659)	(1.170)	(0.535)	(1.073)	(0.623)	(1.160)	(0.598)	(1.173)
	*		**		*		**		**		**		**	
MATHresid1 or	-2.491**	-0.373	-1.807*		-2.396***	-0.056	-1.649*		-2.547***	0.005	-1.740*		-1.588*	
MATHresid2	(0.980)	(1.207)	(0.929)		(0.894)	(1.238)	(0.906)		(0.887)	(1.192)	(0.876)		(0.850)	
	*				*				**					
Need for	-1.446**		-0.992		-0.826				-0.936					
cognition	(0.688)		(0.718)		(0.654)				(0.601)					
Math anxiety					1.788**		2.075***		1.829**		2.149***		2.123***	
					(0.738)		(0.725)		(0.728)		(0.707)		(0.730)	
WINDresid									0.652	-1.080	0.901	-1.061	0.937	-1.077
									(0.823)	(0.653)	(0.863)	(0.643)	(0.814)	(0.677)
									*		*		*	
Log likelihood	-196.830		-198.489		-194.327		-195.684		-193.381		-194.681		R^2=0.397	

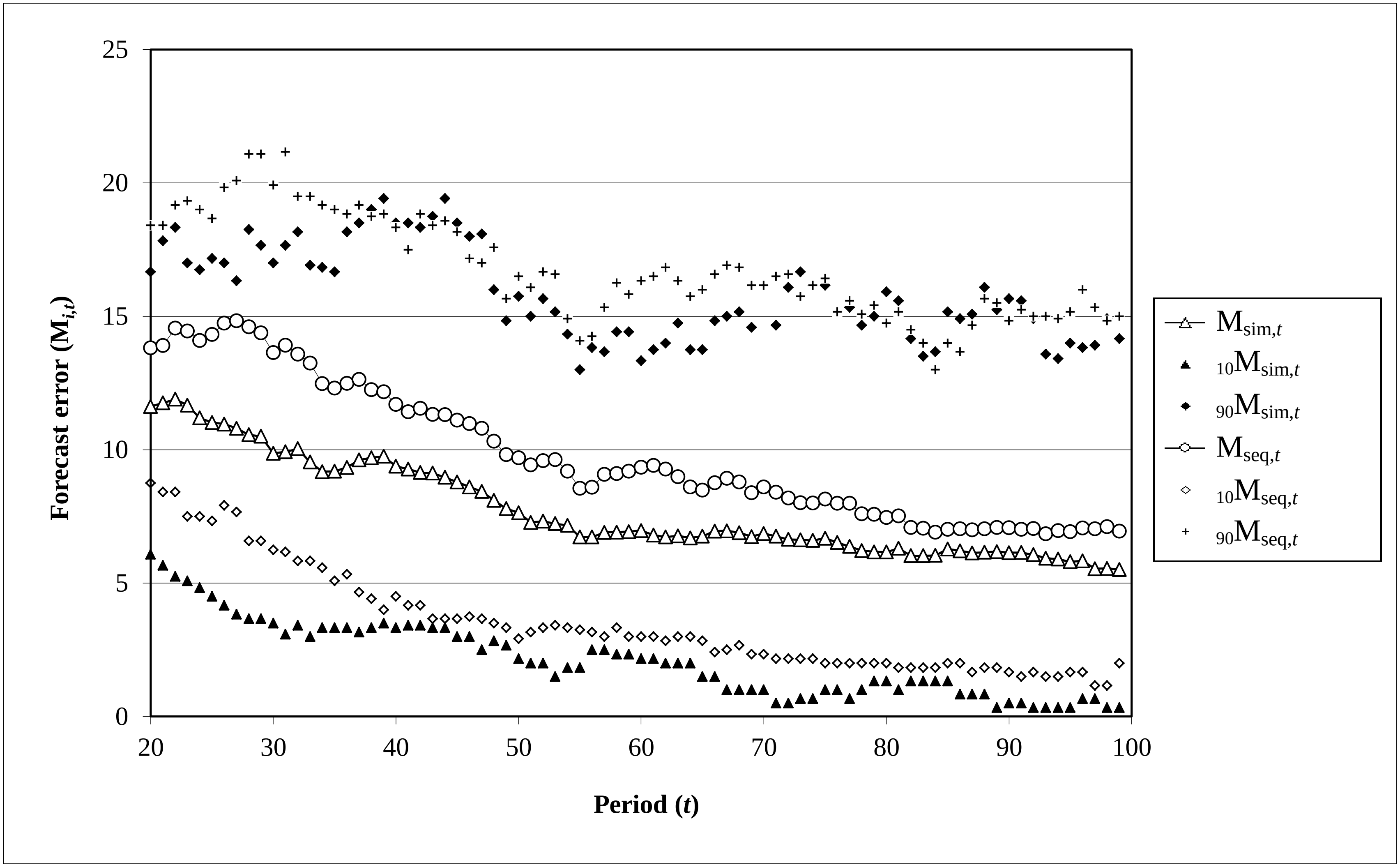
Note: See the note for Table 2.

Table 4: Estimation of forecasting performance on cognitive "residuals", personality and demographic characteristics, and intrinsic forecasting ability

	Model 4c		*Model 4d*		*Model 5c*		*Model 5d*		*Model 6c*		*Model 6d*		*Model 7d*	
REGRESSOR	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}	T_{sim}	T_{seq}
intercept	0.630 (0.864)		0.571 (0.827)		1.013 (0.844)		1.017 (0.800)		0.992 (0.785)		0.984 (0.782)		0.804 (0.773)	
WMresid2	-0.702 (0.658)	-4.500*** (0.950)	-0.735 (0.594)	-4.519*** (0.995)	-0.832 (0.605)	-4.801*** (0.977)	-0.844 (0.574)	-4.879*** (1.032)	-1.471** (0.645)	-4.288*** (1.079)	-1.133* (0.621)	-4.772*** (1.019)	-1.175* (0.616)	-4.451*** (0.931)
(continued)	***		***		***		***		**		***		***	
STMresid2	-1.266* (0.722)	-4.004*** (1.259)	-1.318* (0.743)	-4.005*** (1.235)	-1.304 (0.792)	-4.426*** (1.066)	-1.327 (0.860)	-4.512*** (1.060)	-1.685*** (0.626)	-3.957*** (1.092)	-1.573** (0.755)	-4.599*** (1.052)	-1.612** (0.702)	-4.264*** (1.085)
(continued)	**		**		**		**		**		**		**	
MATHresid2	-1.759** (0.688)		-1.817*** (0.600)		-1.601** (0.688)		-1.633*** (0.614)		-1.769*** (0.631)		-1.705*** (0.591)		-1.550** (0.588)	
Need for cognition	-1.114 (0.762)		-1.370* (0.746)											
Math anxiety					1.936*** (0.646)		2.004*** (0.629)		1.777*** (0.652)		2.082*** (0.615)		2.059*** (0.639)	
WINDresid									0.648 (0.668)	-1.112 (0.718)	1.019 (0.730)	-0.873 (0.712)	1.021 (0.708)	-0.909 (0.728)
(continued)									*		*		*	
EARLYresid1 or EARLYresid2	3.826*** (1.054)		4.388*** (0.939)		3.636*** (0.965)		4.114*** (0.840)		3.701*** (0.916)		4.073*** (0.851)		3.490*** (0.773)	
Log likelihood	-189.428		-186.319		-186.686		-183.829		-182.690		-182.579		R^2=0.572	

Note: See the note for Table 2.

Figure 1: Forecasting performance of the average and the 10th and 90th percentile subjects in each treatment



Note: $M_{sim,t}$ and $M_{seq,t}$ denote the average $M_{i,t}$ in T_{sim} and T_{seq}, respectively. $_{10}M_{sim,t}$ and $_{10}M_{seq,t}$ denote the respective 10th percentiles of $M_{i,t}$. $_{90}M_{sim,t}$ and $_{90}M_{seq,t}$ denote the respective 90th percentiles of $M_{i,t}$.

APPENDIX: Supplementary Design and Implementation Details

Forecasting sequences and strategies

In the Ω_t sequences, only the η_t streams vary across subjects. Their first 75 periods are generated randomly (thereafter, the η_t streams repeat a previous segment due to the matching of the EARLY and LATE periods as well as the eight periods preceding them), subject to the restriction that they are representative in terms of Ω_t-complexity, i.e., the complexity of extracting the seasonal pattern.

In particular, the theoretically most important complexity characteristic is the frequency of events with which subjects encounter the full range of the $\gamma_s+\eta_t$ distributions, conditional on γ_s, since only after observing the range can one determine γ_s with certainty. The arguably most salient aspect of this complexity characteristic is the frequency of events with which the range of a given $\gamma_s+\eta_t$ distribution, conditional on γ_s, can be inferred from successive seasonal realizations of Ω_t and B_t that appear on the screen in a given period. To operationalize this complexity characteristic, all the 75-period η_t streams contain six such events (summed across seasons), six being approximately the sample mean of the frequency of the events for randomly generated 75-period η_t streams.

Another complexity characteristic common to all of the selected 75-period η_t streams is that their sample mean does not significantly differ from zero (at $p<0.01$ using the t-test). Also, the sampling variance of the 75-period η_t streams measured in period 45 varies between 27 and 37. These are approximately the 10th and 90th percentiles, respectively, of the corresponding sampling variance distribution for randomly generated 75-period η_t streams. This condition is to ensure that the η_t streams are not too improbable in the early stages of the task where most learning is expected to occur.

The seemingly most efficient forecasting strategy would first focus on detecting the length of the seasonal pattern, perhaps by experimenting with various lengths, and then on accumulating season-specific information for each of the $\gamma_s+\eta_t$ distributions, conditional on

γ_s, to be able to extract the means of the distributions, γ_s. Nevertheless, a debriefing questionnaire suggests that most subjects relied on less efficient (and likely more memory-intensive) forecasting strategies, attending to *successive* Ω_t-B_t values in an attempt to create a long enough "virtual" sequence of $\gamma_s+\eta_t$ values that would allow them to gradually recognize the seasonal pattern. The debriefing questionnaire also offers suggestive evidence that subjects with higher working memory used more efficient forecasting strategies resembling the efficient strategy described above. This raises the possibility of an indirect "ability-strategy-performance" channel but we do not address the (relative) importance of the channel due to the rather casual evidence on forecasting strategies.

Experimental instructions

In the paper instructions preceding the computerized forecasting task, subjects observe examples of seasonal patterns of various lengths and are advised to attend to the observed past values of Ω_t-B_t = $\gamma_s+\eta_t$ to be able to gradually extract the seasonal parameters, γ_s. Furthermore, before proceeding to the forecasting task, subjects are required to complete computerized training screens that test their understanding of how Ω_t is collectively determined by its three components. However, subjects are told neither how many nor which past values of Ω_t-B_t to attend to.

The detailed task-property feedback in the instructions is meant to further suppress the activation of simplifying heuristics and to instead encourage the use of memory-intensive, financially rewarding forecasting strategies described earlier. The detailed, example-oriented nature of the instructions is further meant to reduce the likelihood that subjects impute their own, possibly erroneous, forecasting context based on their past experience with solving "similar" forecasting problems (in the sense of Harrison and List, 2004).

Timing of forecasting screens

In T_{sim}, subjects observe the two parallel (B_{t+1},…,B_{t-7}) and (Ω_t,…,Ω_{t-7}) screens for 15 seconds each. In T_{seq}, subjects observe the (B_{t+1},…,B_{t-7}) screen for 10 seconds and subsequently the (Ω_t,…,Ω_{t-7}) screen for 15 seconds. While this arrangement does not offer the same total time across treatments for observing the forecast-relevant information, it does offer the same "processing" time of 15 seconds for combining the forecast-relevant information, be it visually in T_{sim} or virtually in T_{seq}. As regards the remaining screens, the feedback screen appears for 5 seconds, and the two screens where subjects place their forecasts and bets are not time-constrained, allowing subjects to go along the forecasting task at their own pace.

Feedback

In each period *t*, subjects are told by how much their forecast, F_{t+1}, is above or below Ω_{t+1}. They are repeatedly reminded in the instructions that η_{t+1} is unpredictable, and they are guided through the implications of the presence of η_{t+1} for their interpretation of the observed "noisy" forecast errors, Ω_{t+1}-F_{t+1}. Judging from responses in a debriefing questionnaire, the instructions were successful in achieving subjects' understanding of the role and implications of η_t, something that people apparently have trouble comprehending in forecasting experiments where the implications of randomness are (often purposefully) not clarified (e.g., Dwyer et al., 1993; Hey, 1994; Maines and Hand, 1996; Stevens and Williams, 2004).

Providing only current-period forecast errors rather than a sequence of past forecast errors is meant to limit the possibility that subjects apply a simplifying feedback-tracking (exponential smoothing) forecasting heuristic often reported in the forecasting literature (e.g., Hey, 1994). We nevertheless note the potential caveat that, due to subjects' varying desire to know more about their forecasting performance progress, not providing more extensive visual feedback might lead to subjects allocating differential amounts of their scarce memory resources to keeping track of how well they are doing, which might in turn dilute the power of working and short-term memory in explaining forecasting performance *per se*. Arguably, however, providing current-period feedback is still better than providing none (e.g., Hey, 1994). Throughout the task, subjects are not provided with earnings feedback (beyond what they can infer from their forecast errors) in order to limit the potential impact of wealth accumulation on forecasting performance.

Betting

To make the betting scheme conceptually transparent, the paper instructions explain in detail that not only forecasting accuracy pays, but also that the more accurately subjects forecast on average, the more profitable betting above 50 ECU becomes on average. As mentioned above, subjects are also guided through the implication of the presence of η_{t+1} for the interpretation of their "noisy" forecast errors, F_{t+1}-Ω_{t+1}. One of the computerized training screens preceding the forecasting task tests subjects' understanding of the payoff function. A full payoff table is provided to subjects but they are reminded that it is far more important to understand the simple logic of how to bet profitably. The instructions also provide subjects with basic context for why they are required to bet on their forecasts in order to make it less likely that subjects provide their own, possibly misleading betting context.

Subject pool and recruitment

Subjects were recruited from Prague Universities, namely the University of Economics, Czech Technical University, Charles University, and Anglo-American College, with the majority of subjects recruited from the first two universities. Czech Technical University is a relatively non-selective university admitting technically-oriented students with heterogeneous educational background, while the University of Economics is more selective university predominantly offering education in economics, management and accounting. We do not detect any differences in forecasting performance related to subjects' field of study, though the sample sizes involved in the comparisons are too small to draw any firm conclusions.

The experiment was run in eleven sessions. Due to concerns that subjects in successive experimental sessions might share information relevant for performing well in the forecasting task as well as some of the cognitive tests, every attempt was made to ensure that successive sessions overlapped or were sufficiently apart from each other, or that subjects in adjacent, non-overlapping sessions were recruited from different universities or university campuses. In retrospect, subjects' behavior in the experiment – especially the lack of "perfect" performance in early stages of the forecasting task – suggests little or no degree of social learning.

A total of 136 subjects completed the experiment, 12 of which were excluded from the analysis. Namely, eight subjects did not meet requirements of the working memory test (their performance on the equation-solving part of the test fell below the required 85% success rate), and four failed to follow our reminder not to make any notes during the forecasting task itself. In the last couple of sessions, we replaced these 12 subjects with new ones in order to obtain the final set of 62 subject pairs facing identical Ω_t sequences across treatment.

Working Paper Series
ISSN 1211-3298
Registration No. (Ministry of Culture): E 19443

Individual researchers, as well as the on-line and printed versions of the CERGE-EI Working Papers (including their dissemination) were supported from the following institutional grants:

- Economic Aspects of EU and EMU Entry [Ekonomické aspekty vstupu do Evropské unie a Evropské měnové unie], No. AVOZ70850503, (2005-2011);
- Economic Impact of European Integration on the Czech Republic [Ekonomické dopady evropské integrace na ČR], No. MSM0021620846, (2005-2011);

Specific research support and/or other grants the researchers/publications benefited from are acknowledged at the beginning of the Paper.

Published by
Charles University in Prague, Center for Economic Research and Graduate Education (CERGE)
and
Economics Institute ASCR, v. v. i. (EI)
CERGE-EI, Politických vězňů 7, 111 21 Prague 1, tel.: +420 224 005 153, Czech Republic.
Printed by CERGE-EI, Prague
Subscription: CERGE-EI homepage: http://www.cerge-ei.cz

Phone: + 420 224 005 153
Email: office@cerge-ei.cz
Web: http://www.cerge-ei.cz

The paper is available online at http://www.cerge-ei.cz/publications/working_papers/.

ISBN 978-80-7343-260-7 (Univerzita Karlova. Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-252-1 (Národohospodářský ústav AV ČR, v. v. i.)

CERGE-EI
P.O.BOX 882
Politických vězňů 7
111 21 Praha 1
Czech Republic
http://www.cerge-ei.cz