When high working memory capacity is and is not beneficial for predicting nonlinear processes

Helen Fischer¹ · Daniel V. Holt¹

Published online: 27 October 2016
© Psychonomic Society, Inc. 2016

Abstract Predicting the development of dynamic processes is vital in many areas of life. Previous findings are inconclusive as to whether higher working memory capacity (WMC) is always associated with using more accurate prediction strategies, or whether higher WMC can also be associated with using overly complex strategies that do not improve accuracy. In this study, participants predicted a range of systematically varied nonlinear processes based on exponential functions where prediction accuracy could or could not be enhanced using well-calibrated rules. Results indicate that higher WMC participants seem to rely more on well-calibrated strategies, leading to more accurate predictions for processes with highly nonlinear trajectories in the prediction region. Predictions of lower WMC participants, in contrast, point toward an increased use of simple exemplar-based prediction strategies, which perform just as well as more complex strategies when the prediction region is approximately linear. These results imply that with respect to predicting dynamic processes, working memory capacity limits are not generally a strength or a weakness, but that this depends on the process to be predicted.

Keywords Prediction · Working memory capacity · Function-learning · Rule-based versus exemplar-based · Nonlinear dynamic processes

Making predictions about the development of dynamic processes is important in many areas of life, be it predicting the

weather, monitoring a patient in critical care, or controlling an industrial power plant. However, research as early as Wagenaar and Sagaria (1975) has shown that the accuracy of predictions varies widely, particularly when processes do not follow a simple linear pattern. In this article, we focus on the contribution of working memory capacity (WMC) for explaining this variation in predictions and its interaction with the type of process to be predicted. Intuitively, higher memory capacity should improve prediction accuracy because people need to consider information about the past of a process in order to forecast its future. And, indeed, research has shown that higher WMC is related to higher prediction accuracy for continuous processes (function learning) and categorization tasks (Bröder, Newell, & Platzner, 2010; Lewandowsky, Yang, Newell, & Kalish, 2012; McDaniel, Cahill, Robbins, & Wiener, 2014). One explanation for this observation is that higher WMC is associated with an improved ability to actively maintain and manipulate information, which is needed to calibrate cognitive prediction algorithms to the learning data and to abstract systematic regularities. These abstracted regularities, or “rules,” can then be used for prediction (McDaniel et al., 2014). However, numerous studies following the *ecological rationality* approach (e.g., Marewski, Gaissmaier, & Gigerenzer, 2010) have shown that in a surprising number of situations simple, information-frugal heuristics—which rely on minimal information and are not working-memory intensive—perform just as well as complex rules (Marewski & Schooler, 2011). Similar arguments have been made for the limited capacity of working memory in general (Cowan, 2010). Rather than being a compromise between processing capacity and metabolic efficiency, a limited WMC may confer genuine advantages. One reason is that because higher WMC endows people with the ability to process more information in parallel, they are more likely to employ complex strategies even when simpler solutions are available (for a review, see

✉ Helen Fischer
helen.fischer@psychologie.uni-heidelberg.de

¹ Department of Psychology, Heidelberg University, Hauptstr. 47-51, 69117 Heidelberg, Germany

Wiley & Jarosz, 2012). In this line of reasoning, capacity limits are a strength that may foster cognitively efficient solution strategies in many situations. Combining these views, higher WMC may be beneficial for predicting processes that require the abstraction of complex rules, but not for processes where simple, information-frugal prediction strategies are sufficient.

To investigate this hypothesis, we used a learning paradigm with different types of exponential processes (MacKinnon & Wearing, 1991). Participants were trained with the beginning of a process and then predicted how the process would continue. Information about each process was presented trial by trial and only in numerical form (see Fig. 1). Participants were given a time point (e.g., 5) and had to predict the value of the process at that time point (e.g., 88). During feedback trials (see Fig. 1, left panel) participants received feedback in form of the correct value of the process (e.g., 208). Then the next time point (e.g., 10) was displayed for the next prediction. During test trials (see Fig. 1, right panel) the procedure was identical, except that no feedback was given. After participants made their prediction and pressed the enter key, the next time point was displayed to make the next prediction. The beginning and end of the training phase consisted of feedback trials with an intermediate block of test trials, which has been shown to enhance transfer of learning (Kang, McDaniel, & Pashler, 2011). Participants did not see results of previous trials, and no summary information (e.g., in the form of graphs) was made available. Instead, participants needed to rely on memory to make predictions. In contrast to typical function learning paradigms (e.g., McDaniel et al., 2014), we were interested in how participants predict the development of ongoing processes rather than learn an abstract functional relation. Data points were therefore presented in sequential order (rather than random), and each data point was presented only once (rather than multiple times).

We used variations of exponential processes for three reasons. First, they are practically relevant, as many processes in real life show exponential dynamics: The spreading of diseases, tipping points in ecological systems, and population growth can all have dramatic consequences if not anticipated early on. Second, empirical research has shown that people find exponential processes difficult to predict (Wagenaar & Sagaria, 1975), but little is known about the effects of individual differences on prediction accuracy. Third, choosing a positive versus negative exponent allows manipulating whether the process is asymptotic (increasingly linear) or accelerating (increasingly nonlinear). Exponential processes comprise both relatively linear regions (e.g., Fig. 2, Panel A, training region) and highly nonlinear regions (e.g., Fig. 2, Panel B, extrapolation region). Wagenaar and Sagaria (1975) described this phenomenon as the cognitively “easy part” and “hard part” of exponentiality. Using a full factorial variation of type (accelerating and decelerating) and direction (increasing and decreasing) for exponential functions allows us to investigate

how well simple linear prediction strategies perform in comparison to more complex strategies in different environments.

Prediction strategies can be summarized in two broad classes: rule based and exemplar based. Rule-based models assume that participants abstract a global rule summarizing the ensemble of information of the process to be predicted (McDaniel & Bussemeyer, 2005). To do so, participants use the feedback provided to update the parameters of their rule in order to calibrate their prediction algorithm to the training data (Koh & Meyer, 1991). Exemplar-based models, in contrast, assume that participants store single exemplars of cue-criterion mappings in memory (Nosofsky, 1988), and that only the most similar training exemplars are retrieved for extrapolation. The *extrapolation association model* (EXAM; DeLosh, Bussemeyer, & McDaniel 1997) and the *population of linear experts* (POLE; Kalish, Lewandowsky, & Kruschke, 2004) assume that participants learn associations between x and y values (EXAM) or between x values and a matching linear function (POLE). Participants extrapolate linearly through the two most similar x and associated y values (EXAM) or use the most similar x value and its expert (POLE). Rule-based and exemplar-based strategies differ with respect to how much training information they use and how well they use it. Rule-based strategies are more information intensive in that a substantial amount of training information is used to induce and calibrate the rule, even if the resulting rule itself is comparatively simple. Exemplar-based strategies, in contrast, are information frugal in that only the most similar training exemplars are used for making a prediction. And only rule-based, but not exemplar-based, strategies imply that participants calibrate their cognitive prediction algorithms to the training data.

If exemplar-based strategies use less training information and are less calibrated, how can they *ever* make as accurate predictions as complex rules? This depends on the type of the process. In asymptotic processes, even if one abstracted the correct function rule, this would not pay off in terms of prediction accuracy because complex rule-based predictions and simple linear predictions based on the most similar exemplars overlap in this case (see Fig. 2, bottom panel). Moreover, rule-induction tends to be error prone because several problem-solving steps need to be performed in working memory (Beilock & DeCaro, 2007), whereas simple linear extrapolations are typically performed with near optimal accuracy (Bussemeyer, Byun, DeLosh, & McDaniel, 1997). Abstracting information-intensive rules hence cannot outperform simple linear strategies for predicting the later parts of asymptotic processes and may even impair prediction. For increasingly nonlinear processes such as accelerating functions (see Fig. 1, top panel), in contrast, rule-based strategies that capture the process’ trend should be more accurate because simple linear predictions would lead to an underestimation of the trajectory.

Fig. 1 Procedure of the experiment showing feedback trials (left panel) and test trials (right panel). Participants were given a time point and entered the predicted value of the process at that time point. During

training trials participants were given feedback in the form of the correct value of the process, then the next time point was displayed for the next prediction. During test trials no feedback was given.

It seems plausible that higher WMC should lead to more accurate predictions in processes that benefit from the information-intensive induction of rules, that is, the accelerating processes in this experiment. Rule-induction requires memorizing cue and criterion values, estimating their differences, and updating the rule accordingly. These processes arguably encompass storage and transformation of information—key facets of working memory (Oberauer, Süß, Wilhelm, & Wittman, 2003). In categorization and multiple-cue judgment, WMC indeed predicted performance in tasks requiring rule-based strategies (Hoffmann, von Helversen, & Rieskamp, 2014). The involvement of WMC

is particularly high when more cues need to be considered or more complex rules need to be abstracted (Karlsson, Juslin, & Olsson 2008; Mata et al., 2012). WMC is also associated with accuracy in a prototypical rule-induction task—Raven’s Progressive Matrices— particularly for items requiring complex rules (D. R. Little, Lewandowsky, & Craig, 2014; Wiley, Jarosz, Cushen, & Colflesh, 2011). Despite these compelling connections, there is only one study directly assessing how WMC affects the prediction of continuous processes: McDaniel et al. (2014) suspected that higher WMC allows participants to actively maintain a range of cue-criterion values and to concurrently compare them over trials to abstract

Fig. 2 Processes used in the experiment: (A) accelerating increasing, (B) accelerating decreasing, (C) asymptotic decreasing, and (D) asymptotic increasing. Participants learned about the processes from time points 0 to 85 (training), and predicted the processes from time points 90 to 115

(extrapolation). Participants predicted exponential processes (black), and their respective quadratic twins (gray) that run through the same two training exemplars closest to extrapolation.

a functional rule. Higher WMC participants were indeed more likely to use rule-based as opposed to exemplar-based strategies and achieved higher prediction accuracy. In this study, however, only a single type of process (V-shaped) was used, which benefits from the application of rule-based strategies.

In cases where simple linear prediction models should perform optimal—asymptotic processes in the present case—higher WMC might not pay off in terms of better predictions. Exemplar-based predictions rely on recall of only the most similar training exemplars. Given that in our experimental design these are the most recently completed trials before extrapolation, this should require very little working memory resources. Supporting this assumption, Hoffmann et al. (2014) found that WMC was not related to performance in exemplar-based judgment tasks at all. It is unclear, however, whether higher WMC participants will use simple exemplar-based strategies if this is sufficient, or whether they will try to apply rules even if this strategy does not pay off. Some studies found that higher cognitive capacity fosters adaptive strategy use (Bröder, 2003; Lewandowsky et al., 2012). Therefore, one possibility in the case of asymptotic processes is that higher WMC participants may use the simpler exemplar-based strategies. Other studies, however, found higher WMC to be detrimental for performance in tasks where simple strategies are optimal. In the classic water jug problem, a complex solution strategy is needed for the first trials, but a simpler strategy becomes available later on. Higher WMC participants were less likely to switch to the simple strategy and instead persisted in using the overly complex strategy (Beilock & DeCaro, 2007). Similarly, in a category learning task in which complex hypotheses testing effectively impeded learning, higher WMC participants performed worse than lower WMC participants (DeCaro, Thomas, & Beilock, 2008). In a judgment task for which a simple similarity-based strategy performed optimally, participants were more likely to use the optimal strategy when fewer WM resources were available under cognitive load (Hoffmann, von Helversen, & Rieskamp, 2013). Although people generally tend to search for rules in sequences, even if no pattern exists (Wolford, Newman, Miller, & Wig, 2004), this tendency may be particularly pronounced for people with higher cognitive capacity.

Given these findings, we expect the following effects of higher working memory capacity: Higher WMC should improve accuracy when predicting complex, accelerating processes that require rule induction. Higher WMC may not pay off, however, when predicting simple, asymptotic processes. Instead, higher WMC participants are more likely to use complex rule-based approaches, even in cases where simple exemplar-based strategies suffice. We designed a prediction experiment based on the exponential functions shown in Figure 2 to test this assumption and several hypotheses derived from it.

First, we expect an interaction of WMC and stimulus type with respect to extrapolation accuracy. The increased use of

rule-based prediction by high WMC participants should improve prediction accuracy for the increasingly nonlinear part of accelerating functions. However, as complex rules have no benefit for predicting the increasingly linear part of asymptotic processes, we expect no difference between higher and lower WMC in these conditions.

Second, we expect participants with higher WMC to show higher accuracy during the training phase, as they are better able to induce a rule adequately describing the training information—that is, a rule that is better calibrated to the training data.

Third, we expect lower WMC participants to rely more on the most recent information available for extrapolation, while higher WMC participants should integrate more training information. To test this hypothesis, we constructed a quadratic “twin” process for each exponential stimulus (see Fig. 2). For each point in the training phase, exponential processes and their quadratic twins had a different value, except for the two points just before the extrapolation region, which were identical. For each exponential stimulus, participants completed the same training and prediction procedure with the corresponding quadratic twin process. Hence, the more information (beyond the two final points) participants use for making a prediction, the more their predictions should differentiate between the two processes of a pair. The effect should be conspicuous for accelerating processes but weak or absent for asymptotic processes because their trajectory in the extrapolation region can be accurately approximated by simple linear extrapolation. This design extends previous research by McDaniel et al. (2014), who observed that rule-based versus exemplar-based learners did not just differ in accuracy but also showed different patterns of predictions.

Finally, as previous studies did not control for potential effects of explicitly instructing participants to search for a rule to make predictions, one group of participants was instructed to use the feedback provided in order to find a rule describing each process. The other group was instructed to simply rely on their intuition to make predictions. We hypothesized that this may have effects comparable to the WMC-related differences in rule use described above.

Method

Following the *statement of transparency* proposed by Simmons, Nelson, & Simonsohn (2012), we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Design Each participant predicted two exponential processes and their respective quadratic twins. Type of the exponential processes (accelerating or asymptotic) and instruction (rule search or intuitive) were varied between

subjects, direction of the process (increasing or decreasing) within subjects. WMC tests were interleaved between prediction tasks.

Participants In total, 295 students from Heidelberg University participated in the study, providing 290 complete data sets (201 female, mean age = 23.0 years, $SD = 2.7$). We decided in advance of data collection to aim for 300 participants to achieve sufficient statistical power for detecting small to medium effects. All participants gave written informed consent and were rewarded with 5 Euros.

Materials Four exponential and quadratic function pairs were constructed such that the twinned functions were identical at the two points closest to extrapolation (time points 80 and 85) but different at all other points (see Fig. 1 and Table 1). Functions were constructed for a time range (x value) from 0 to 115, increasing in steps of 5 for a total of 24 time points.

Working memory tests WMC was assessed using two computer-based tasks: digit span backward (DSB) and letter–number sequencing (LNS; see Wechsler, 2008). We selected these WMC measures as they are easy to administer yet are closely related to more complex measures often used in working memory research, such as operation span (Shelton, Elliott, Hill, Calamia, & Gouvier, 2009). For DSB, increasingly longer series of digits were displayed, which participants had to repeat in reverse order using an on-screen keyboard. In the LNS task, increasingly longer series of alternating digits and numbers were presented. Participants had to sort the numbers in ascending and the letters in alphabetical order before responding. Span lengths ranged from four to a maximum of eight for DSB and nine for LNS with two (DSB) or three (LNS) trials per span length. Tests were finished after failing all trials of a span length. Individual WMC was computed as the mean of the z -standardized scores for both subtests. Data for the digit span forward task were also collected, but excluded from the analysis as it was the only task that did not involve active manipulation of information.

Procedure Participants were instructed that they had to predict four different processes over time with the help of

feedback. The nature of the processes was left unspecified. One group of participants was instructed to “use the feedback provided to derive a rule describing the trajectory of the process” and to “predict the value of the process at a given time point as accurately as possible.” The other group was instructed to “use the feedback provided to gain an intuition about the trajectory of the process” and to “use the intuition that you gained about the trajectory of the process to make a prediction about the value of the process at a given time point as accurately as possible.”

As shown in Figure 1, during each trial, participants were shown the current time point (labeled *Time*) approximately in the middle of the computer screen and entered their prediction for that time point as a number in a textbox (labeled *Prediction*) directly below. Time points then increased in steps of 5 from 0 to 115. For each time point, participants made one prediction. During time points 0 to 25 and 60 to 85, participants received feedback showing the correct value of the process below the time point and their last prediction (labeled *Correct*). During time points 30 to 55 in the training phase, no feedback was given before the next time point was displayed. This block of trials without feedback served to enhance the learning process through an intermediate testing phase, as previous research found that training-test conditions constitute a more effective learning environment than training-only conditions (Kang et al., 2011). The final block of trials without feedback (time points 90 to 115) constituted the main extrapolation region (see Fig. 2). In the remainder of this text, “training” refers to the last block of training trials (time points 60 to 85) because this block contained the twin process manipulation, and “prediction” refers to the final extrapolation phase (time points 90 to 115).

Results

Outliers Single predictions more than five standard deviations above or below the mean of a time point were excluded from calculating individual averages per process. If less than half of predictions remained for a process, the process was treated as missing for this participant (2.3 % of processes). Furthermore, 35 participants received incorrect feedback not matching the training function for the first half of two processes (asymptotic

Table 1 Stimulus functions

Type	Direction	Exponential function	Quadratic twin function
Accelerating	Increasing	$3,000 + e^{(0.045 * x + 4.2)}$	$0.6494 * (x + 12.355)^2 - 99.13$
Accelerating	Decreasing	$2,000 + e^{(-0.045 * x + 8)}$	$0.2445 * (x - 89.042)^2 + 2,061$
Asymptotic	Increasing	$1,500 - e^{(-0.045 * x + 7)}$	$-0.0108 * (x - 137.635)^2 + 1,506.15$
Asymptotic	Decreasing	$1,500 - e^{(0.045 * x + 2)}$	$-0.0491 * (x + 57.979)^2 + 2,165.11$

decreasing quadratic and asymptotic increasing exponential). As the present analysis is based only on responses in the second half of each process, we retained these participants in the analysis.

Dependent variables As an index of accuracy we used the root mean relative error (RMRE). To correct for different scaling, the relative error for each prediction is determined by dividing the absolute deviation of the prediction from the correct function value at each point by the correct function value. Then the mean of relative errors is calculated for predictions of each process, and a square root transformation is applied to reduce skew. To assess how much participants differentiated predictions between the twinned processes, we calculated the difference of mean predictions for each pair of twinned processes divided by the standard deviation of differences.

Effects of instruction, process type, and direction on accuracy We applied a linear model with instruction, process type, and process direction as factors; WMC as continuous predictor; and RMRE as dependent variable. The dedicated rule-search versus intuitive instructions did not exert an effect on prediction accuracy, $F(1, 270) = .39, p = .53$. We therefore excluded this factor from further analyses. There were main effects of both process type, $F(1, 270) = 1674.36, p < .001, \eta^2 = .86$, and direction $F(1, 270) = 356.67, p < .001, \eta^2 = .57$, and a significant interaction of the two, $F(1, 270) = 263.42, p < .001, \eta^2 = .49$. Prediction errors were higher for accelerating compared to asymptotic processes and higher for decreasing compared to increasing processes. However, prediction errors were disproportionately high for the accelerating decreasing process (see Table 2 for descriptive statistics).

Effects of WMC on accuracy In line with our expectation, we found an interaction of WMC and process type with respect to prediction accuracy, $F(1, 270) = 6.67, p = .01, \eta^2 = .02$, but no main effect of WMC, $F(1, 270) = 2.13, p = .15$, and no interaction with direction, $F(1, 270) = 0.67, p = .42$. Specifically, high WMC was associated with lower RMREs in the accelerating conditions, $r(161) = -.18, p = .02$, but not in

the asymptotic conditions, $r(121) = -.03, p = .77$. We further investigated whether WMC affected accuracy during training as a sign of rule induction. There was an overall positive effect of WMC on training accuracy, $F(1, 274) = 13.45, p < .001, \eta^2 = .05$, no statistically significant interaction of WMC and process type (albeit close to significance), $F(1, 274) = 3.54, p = .06, \eta^2 = .01$, and no interaction with direction, $F(1, 274) = .90, p = .34$. High WMC was associated with lower training error in both accelerating, $r(160) = -.19, p = .02$, and asymptotic conditions, $r(125) = -.25, p < .01$.

Differentiation To investigate the amount of training information that participants integrate into their prediction strategies, we calculated the difference in predictions between exponential and quadratic process twins. A differentiation score different from zero means that participants' predictions cannot be explained by simple linear extrapolation based on only the final two training points. The differentiation score was calculated such that positive values indicate a more extreme extrapolation of exponentially accelerating processes relative to their quadratic twins or a less extreme extrapolation of asymptotic processes. To allow summarizing over different numerical ranges, differentiation scores were divided by their respective standard deviations. In contrast to our expectation, differentiation was not related to WMC overall, $F(1, 265) = .08, p = .78$, and there was no interaction of WMC with process type, $F(1, 265) = .52, p = .47$, but with process direction, $F(1, 265) = 4.61, p = .03, \eta^2 = .02$. Differentiation was only correlated with WMC in the accelerating decreasing condition, $r(153) = .18, p = .02$, but not in the other conditions, $ps > .07$.

Discussion

This study investigated how working memory capacity limits are associated with accuracy and strategy use in predicting nonlinear processes. Although it may seem intuitive that higher WMC should always be associated with better predictions, we found an interaction between WMC and the type of the process to be predicted: For both increasing and

Table 2 Prediction accuracies and differentiation for the four exponential processes

Process		Prediction error (RMRE)	Training error (RMRE)	Differentiation Exponential-Quadratic
Type	Direction			
Accelerating	Increasing	.43 (.11)	.23 (.07)	.01
Accelerating	Decreasing	.83 (.21)	.15 (.03)	-.12
Asymptotic	Increasing	.09 (.06)	.25 (.10)	.02
Asymptotic	Decreasing	.13 (.11)	.28 (.11)	.20

Note. Means with standard deviations in parentheses. Differentiation $SD = 1$.

decreasing processes, higher WMC participants were more accurate predicting the accelerating processes but not more accurate predicting the asymptotic processes. Crucially, this interaction in prediction accuracy appeared *despite* higher WMC participants' better calibration to training data in both accelerating and asymptotic processes. Thus, higher WMC appears to be associated with better calibrated strategies that pay off when predicting complex accelerating processes but do not pay off when predicting simple asymptotic processes.

To measure strategy use, we employed two operationalizations (calibration to training data and differentiation between process twins) independent of prediction accuracy. We distinguish two main types of prediction strategy: Simple linear extrapolation based on the two most recent training exemplars only and more complex strategies requiring information encountered prior to these two exemplars. We treat the former as a case of exemplar-based prediction, such as described by EXAM or POLE (DeLosh et al., 1997; Kalish et al., 2004), and the latter as evidence for the use of more complex, rule-based strategies. This interpretation is based on the assumption that memory for the two most recent training exemplars is nearly perfect and that there is no (or at best a negligible) effect of earlier training exemplars when using simple linear extrapolation. Even for participants with low WMC, it seems a reasonable assumption that the two most recently presented and most similar exemplars (or linear experts in POLE) are highly available.

WMC was positively correlated with calibration during training, which is considered a prototypical sign of rule induction (Koh & Meyer, 1991). Importantly, calibration increased with WMC both when predicting accelerating processes (where this is the optimal strategy) and when predicting asymptotic processes (where a simpler and equally accurate strategy is available). Apparently, higher WMC participants were generally better able to make use of the learning phase provided to align their prediction strategies to the structure of the training data. This is a particularly noteworthy result because it demonstrates that higher WMC participants did not make better predictions for the asymptotic processes, even though they did use better calibrated strategies during training.

Unexpectedly, the extent to which participants' predictions differentiated between quadratic and exponential process twins revealed an interaction effect of WMC and process direction. WMC and differentiation were significantly correlated only for the accelerating decreasing process. Applying rules that integrate more training information than just the two most recent training exemplars should have produced differentiation for both accelerating processes. A plausible explanation for this discrepancy is that the accelerating decreasing process deviated most from a cognitive default of positive linearity (Busemeyer et al., 1997). It also was the most difficult of all processes to predict, as indicated by the large prediction errors and the high interindividual variability.

The effect of WMC may therefore have been moderated by this difference in task difficulty with a larger benefit of WMC in the more difficult task.

These results suggest that higher WMC does not always lead to more accurate predictions, but does so only if the task is sufficiently difficult and the structure of the process to be predicted benefits from applying complex rules. This is an important qualification of previous findings on how individual WMC affects the accuracy of predictions that show that higher WMC generally leads to increased use of rules and to better predictions (McDaniel et al., 2014). Extending the approach of McDaniel, this study employed the twin-process manipulation to selectively assess the amount of learning data incorporated into prediction strategies. The results show that higher WMC may in fact not provide a prediction advantage for processes where simple exemplar-based strategies perform optimally, but that higher WMC only provides an advantage when rule induction is a beneficial strategy and the task is sufficiently complex. This result is only partially in line with recent work in the area of categorization. Although a benefit of WMC for learning category structure is generally found, stable individual preferences for strategy use seem mostly unrelated to cognitive abilities such as WMC or reasoning (Craig & Lewandowsky, 2012; J. L. Little & McDaniel, 2015). One important difference between typical categorization tasks and most function learning or process prediction paradigms may be that the latter strongly suggest the existence of an underlying numerical rule. Assuming that participants with high WMC are strategically more flexible, they may simply be more likely to use strategies matching the perceived demands of the task.

These results may be comparable to the phenomena of over- and underfitting in machine learning (Hawkins, 2004; Todd & Gigerenzer, 2000). Underfitting occurs when rendering a prediction algorithm more complex or improving its calibration to the training data would pay off in terms of better predictions. Overfitting, in contrast, occurs when higher complexity and better calibration no longer pay off—that is, when a complex compared to a simpler prediction algorithm does not produce more accurate predictions. In this study, both phenomena may have occurred: In the (complex) accelerating processes, lower WMC participants' prediction strategies were less calibrated, and this resulted in worse predictions. In the asymptotic processes, however, higher WMC participants' better calibrated prediction strategies did not produce better predictions compared to lower WMC participants' simpler strategies. Thus, lower WMC participants' strategies underfitted the accelerating processes, and higher WMC participants' strategies resulted in an inefficient overfitting of the asymptotic processes.

Perhaps counterintuitively, higher WMC participants may have used suboptimal strategies particularly when performing the comparatively simple task of predicting asymptotic

processes. What may explain this finding is that predicting asymptotic processes is not only comparatively simple but actually *surprisingly* simple. The steep training region suggests structural complexity (and hence the use of more complex rules), but the extrapolation region runs relatively flat (and hence is best predicted with simple linear strategies) – the cognitively “easy part” and “hard part” of exponentiality described by Wagenaar and Sagaria (1975). Our results hence suggest that in cases where “easy parts” directly follow “hard parts,” higher WMC may not be a helpful resource. On the contrary, higher WMC may even impede switching to simpler approaches. Analogous to Beilock and DeCaro (2007), who have shown that the tendency to inflexibly persist in using more complex approaches is especially pronounced for higher WMC participants, we found that higher WMC participants were prone to overlook simple yet accurate prediction strategies.

This may also explain the seemingly contradictory finding that higher cognitive capacity sometimes leads to more adaptive strategy use (Bröder, 2003; Lewandowsky et al., 2012), but at other times to the use of overly complex, nonadaptive strategies (Beilock & DeCaro, 2007). In the experiment by Beilock and DeCaro (2007) and in the asymptotic processes in this study, complex strategies were successful during earlier trials, but a simpler solution became available later on. It was then that higher WMC participants persisted in using their more complex strategy. Higher WMC participants seem to be able to construct an initially suitable, complex strategy for a given task, but may have difficulty giving up this strategy when a simpler strategy becomes available.

Interestingly, the manipulation of instructions introduced to elicit the use of explicit rule-based or implicit intuitive prediction strategies had no notable effect on performance, thereby supporting previous studies that did not differentiate between explicit and intuitive instructions (e.g., Kang et al., 2011). The lack of a difference may either imply that the instruction manipulation was not strong enough or that the prediction strategies participants use are relatively robust and little affected by instructions in general. Supporting the second interpretation, similar instructions used in a dynamic systems control study exerted only a small effect on performance (Hundertmark, Holt, Fischer, Said, & Fischer, 2015). For future research, an alternative approach would be a cognitive load manipulation to investigate how this affects the use of explicit and implicit prediction strategies.

Cowan (2010) suspected that the positions of working memory capacity limits as a weakness versus as a strength may not be incompatible, but that each one may have its merits. With respect to predicting nonlinear dynamic processes, we found that whether or not higher capacity is beneficial depends on the type of process to be predicted. Higher working memory capacity seems to be associated with the ability to use better calibrated strategies that are more aligned with the

structure of the processes to be predicted. This leads to better predictions of difficult, nonlinear processes, but can result in overfitting and applying overly complex strategies when predicting simple processes. In a word: Finding structure is good, but finding more structure is not necessarily better.

References

- Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 983–998. doi:10.1037/0278-7393.33.6.983
- Bröder, A. (2003). Decision making with the “adaptive toolbox”: Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 611–625. doi:10.1037/0278-7393.29.4.611
- Bröder, A., Newell, B. R., & Platzner, C. (2010). Cue integration vs. exemplar-based reasoning in multi-attribute decisions from memory: A matter of cue representation. *Judgment and Decision Making*, 5, 326–338.
- Busemeyer, J. R., Byun, E., Delosh, E. L., & McDaniel, M. A. (1997). Learning functional relations based on experience with input-output pairs by humans and artificial neural networks. In K. Lamerts & D. Shanks (Eds.), *Concepts and categories* (pp. 405–437). Cambridge, MA: MIT Press.
- Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, 19, 51–57. doi:10.1177/0963721409359277
- Craig, S., & Lewandowsky, S. (2012). Whichever way you choose to categorize, working memory helps you learn. *The Quarterly Journal of Experimental Psychology*, 65, 439–464. doi:10.1080/17470218.2011.608854
- DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition*, 107, 284–294. doi:10.1016/j.cognition.2007.07.001
- DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 968–986. doi:10.1037/0278-7393.23.4.968
- Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, 44, 1–12. doi:10.1021/ci0342472
- Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013). Deliberation’s blindsight: How cognitive load can improve judgments. *Psychological Science*, 24, 869–879. doi:10.1177/0956797612463581
- Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2014). Pillars of judgment: How memory abilities affect performance in rule-based and exemplar-based judgments. *Journal of Experimental Psychology: General*, 143, 2242–2261. doi:10.1037/a0037989
- Hundertmark, J., Holt, D. V., Fischer, A., Said, N., & Fischer, H. (2015). System structure and cognitive ability as predictors of performance in dynamic system control tasks. *Journal of Dynamic Decision Making*, 1, 5. doi:10.11588/jddm.2015.1.26416
- Kalish, M. L., Lewandowsky, S., & Kruschke, J. K. (2004). Population of linear experts: Knowledge partitioning and function learning. *Psychological Review*, 111, 1072–1099. doi:10.1037/0033-295X.111.4.1072

- Kang, S. H., McDaniel, M. A., & Pashler, H. (2011). Effects of testing on learning of functions. *Psychonomic Bulletin & Review*, 18, 998–1005. doi:[10.3758/s13423-011-0113-x](https://doi.org/10.3758/s13423-011-0113-x)
- Karlsson, L., Juslin, P., & Olsson, H. (2008). Exemplar-based inference in multi-attribute decision making: Contingent, not automatic, strategy shifts. *Judgment and Decision Making*, 3, 244–260.
- Koh, K., & Meyer, D. E. (1991). Function learning: Induction of continuous stimulus-response relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 811–836. doi:[10.1037/0278-7393.17.5.811](https://doi.org/10.1037/0278-7393.17.5.811)
- Lewandowsky, S., Yang, L.-X., Newell, B. R., & Kalish, M. L. (2012). Working memory does not dissociate between different perceptual categorization tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 881–904. doi:[10.1037/a0027298](https://doi.org/10.1037/a0027298)
- Little, D. R., Lewandowsky, S., & Craig, S. (2014). Working memory capacity and fluid abilities: The more difficult the item, the more more is better. *Frontiers in Psychology*, 5, 239. doi:[10.3389/fpsyg.2014.00239](https://doi.org/10.3389/fpsyg.2014.00239)
- Little, J. L., & McDaniel, M. A. (2015). Individual differences in category learning: Memorization versus rule abstraction. *Memory & Cognition*, 43, 283–297. doi:[10.3389/fpsyg.2014.00239](https://doi.org/10.3389/fpsyg.2014.00239)
- MacKinnon, A. J., & Wearing, A. J. (1991). Feedback and the forecasting of exponential change. *Acta Psychologica*, 76, 177–191. doi:[10.1016/0001-6918\(91\)90045-2](https://doi.org/10.1016/0001-6918(91)90045-2)
- Marewski, J. N., Gaissmaier, W., & Gigerenzer, G. (2010). Good judgments do not require complex cognition. *Cognitive Processing*, 11, 103–121. doi:[10.1007/s10339-009-0337-0](https://doi.org/10.1007/s10339-009-0337-0)
- Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: an ecological model of strategy selection. *Psychological Review*, 118, 393–437. doi:[10.1037/a0024143](https://doi.org/10.1037/a0024143)
- Mata, R., Pachur, T., Von Helversen, B., Hertwig, R., Rieskamp, J., & Schooler, L. (2012). Ecological rationality: a framework for understanding and aiding the aging decision maker. *Frontiers in Neuroscience*, 6, (19). doi:[10.3389/fnins.2012.00019](https://doi.org/10.3389/fnins.2012.00019)
- McDaniel, M. A., & Busemeyer, J. R. (2005). The conceptual basis of function learning and extrapolation: Comparison of rule-based and associative-based models. *Psychonomic Bulletin & Review*, 12, 24–42. doi:[10.3758/BF03196347](https://doi.org/10.3758/BF03196347)
- McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: Stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General*, 143, 668–693. doi:[10.1037/a0032963](https://doi.org/10.1037/a0032963)
- Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 14, 700–708. doi:[10.1037/0278-7393.14.4.700](https://doi.org/10.1037/0278-7393.14.4.700)
- Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, 31, 167–193. doi:[10.1016/S0160-2896\(02\)00115-0](https://doi.org/10.1016/S0160-2896(02)00115-0)
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution. doi:[10.2139/ssrn.2160588](https://doi.org/10.2139/ssrn.2160588)
- Shelton, J. T., Elliott, E. M., Hill, B. D., Calamia, M. R., & Gouvier, W. D. (2009). A comparison of laboratory and clinical working memory tests and their prediction of fluid intelligence. *Intelligence*, 37, (3), 283–293. doi:[10.1016/j.intell.2008.11.005](https://doi.org/10.1016/j.intell.2008.11.005)
- Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, 23, 727–741. doi:[10.1017/S0140525X00003447](https://doi.org/10.1017/S0140525X00003447)
- Wagenaar, W. A., & Sagaria, S. D. (1975). Misperception of exponential growth. *Perception & Psychophysics*, 18, 416–422. doi:[10.3758/BF03204114](https://doi.org/10.3758/BF03204114)
- Wechsler, D. (2008). *Wechsler adult intelligence scale* (4th ed.). San Antonio, TX: Pearson.
- Wiley, J., & Jarosz, A. F. (2012). Working memory capacity, attentional focus, and problem solving. *Current Directions in Psychological Science*, 21, 258–262. doi:[10.1177/0963721412447622](https://doi.org/10.1177/0963721412447622)
- Wiley, J., Jarosz, A. F., Cushen, P. J., & Colflesh, G. J. (2011). New rule use drives the relation between working memory capacity and Raven's Advanced Progressive Matrices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 256–263. doi:[10.1037/a0021613](https://doi.org/10.1037/a0021613)
- Wolford, G., Newman, S. E., Miller, M. B., & Wig, G. S. (2004). Searching for patterns in random sequences. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 58, 221. doi:[10.1037/h0087446](https://doi.org/10.1037/h0087446)