Archival Report

Biological Psychiatry

Incentives Boost Model-Based Control Across a Range of Severity on Several Psychiatric Constructs

Edward H. Patzelt, Wouter Kool, Alexander J. Millner, and Samuel J. Gershman

ABSTRACT

BACKGROUND: Human decision making exhibits a mixture of model-based and model-free control. Recent evidence indicates that arbitration between these two modes of control ("metacontrol") is based on their relative costs and benefits. While model-based control may increase accuracy, it requires greater computational resources, so people invoke model-based control only when potential rewards exceed those of model-free control. We used a sequential decision task, while concurrently manipulating performance incentives, to ask if symptoms and traits of psychopathology decrease or increase model-based control in response to incentives.

METHODS: We recruited a nonpatient population of 839 online participants using Amazon Mechanical Turk who completed transdiagnostic self-report measures encompassing symptoms, traits, and factors. We fit a dual-controller reinforcement learning model and obtained a computational measure of model-based control separately for small incentives and large incentives.

RESULTS: None of the constructs were related to a failure of large incentives to boost model-based control. In fact, for the sensation seeking trait and anxious-depression factor, higher scores were associated with a larger incentive effect, whereby greater levels of these constructs were associated with larger increases in model-based control. Many constructs showed decreases in model-based control as a function of severity, but a social withdrawal factor was positively correlated; alcohol use and social anxiety were unrelated to model-based control.

CONCLUSIONS: Our results demonstrate that model-based control can reliably be improved independent of construct severity for most measures. This suggests that incentives may be a useful intervention for boosting model-based control across a range of symptom and trait severity.

Keywords: Computational psychiatry, Habits and goals, Incentives, Model-based control, Psychiatric constructs, Reinforcement learning

https://doi.org/10.1016/j.biopsych.2018.06.018

Decisions are sometimes the product of habit, and sometimes the product of planning. The two forms of decision making embody complementary strengths and weaknesses: habits are inflexible (and hence sometimes inaccurate) but require minimal cognitive effort, whereas plans support flexible goal pursuit but require greater cognitive effort. This dichotomy has significant clinical implications, because some psychopathological behaviors can be understood as arising from the hegemony of habits over plans. For example, drug addiction and obsessive-compulsive disorder are associated with an inability to overcome maladaptive habits (1).

Computational models have led to a formal description of the habit-planning dichotomy by specifying precise algorithmic hypotheses (2–4), with wide-ranging ramifications for the neurobiology of decision making and its breakdown in psychopathology. In particular, reinforcement learning models have operationalized habits in terms of "model-free" control and planning in terms of "model-based" control. Model-free control selects actions based on the degree to which they have been rewarded, using cached reward predictions that are updated by trial and error. Because these cached predictions can only be updated by interaction with the environment [though see Gershman *et al.* (5)], they will be insensitive to changes in the environment that have not been directly experienced, leading to the brittleness characteristic of habits. Model-based control selects actions based on an internal model of the environment, which specifies how actions affect the state of the environment. By using mental simulation of the internal model, model-based control can generate sequential plans that adapt flexibly to changes in the environment, even without direct experience. However, mental simulation is cognitively costly, and thus people will prefer model-free control when cognitive resources are scarce (6), accuracy is poorly incentivized (7), or mental simulation produces unreliable reward predictions (2,8).

In the current study, we examine the balance of model-free and model-based control using a reinforcement learning paradigm that allows us to quantify the degree to which

ISSN: 0006-3223

model-based control can be incentivized. This allows us to ask a critical question for the treatment of certain psychiatric symptoms and traits: can incentives ameliorate model-based deficits associated with certain clinical symptoms and traits? We approach this question by collecting a broad range of commonly used self-report measures and well-defined transdiagnostic traits in a diverse online population, measuring the relationship between these measures and a computationally derived estimate of model-based control under different incentive conditions.

Model-Based Control and Psychopathology

Empirical studies indicate that model-based control is disrupted across an array of disorders and psychopathological constructs. In particular, model-based impairments have been revealed in schizophrenia (9), binge eating disorder (10), obsessive-compulsive disorder (10), and methamphetamine dependence (10). However, studies of alcohol dependence have been equivocal, with research finding behavioral deficits (11) and reduced prefrontal signatures of model-based control (12) in detoxified alcohol-dependent patients, but no association between model-based control and problematic drinking in adolescents (13).

While these studies focused on DSM diagnoses, a recent study by Gillan *et al.* (14) examined model-based control among several transdiagnostic traits. They found that a factor containing compulsive behavior and intrusive thought items was negatively associated with model-based control, whereas an anxious-depression factor showed no relationship, and a social withdrawal factor had a small positive association with model-based control (14). On the one hand, comparing diagnostic groups has several problems, including high levels of comorbidity and shared symptoms among people in different diagnostic groups (15–17). On the other hand, using continuous outcomes allows researchers to understand how cognitive processes, such as model-based control, covary with increasing symptom and trait severity, across several clinically related constructs.

Gillan *et al.* (14) measured model-based control using a sequential decision task developed by Daw *et al.* (18), which we refer to as the Daw two-step task. This task has been the standard assessment tool for measuring model-based control in clinical studies. However, recent research has shown that the Daw two-step task does not incentivize model-based control, because model-free and model-based control lead to roughly equivalent performance on the task (19). Moreover, increasing incentives in the Daw two-step task does not increase model-based control (7). For this reason, the task is not useful for addressing the question of whether psychiatric symptoms and traits are associated with an inability to boost model-based control in response to incentives. Fortunately, Kool *et al.* (19) have developed a novel sequential decision task in which model-based control does lead to improved performance, and incentives are effective at amplifying model-based control on the task (7). In particular, informing participants that their earnings would be multiplied by five on particular trials (the "high-stakes" condition) was effective at increasing their reliance on model-based control relative to a baseline ("low stakes") condition. Our goal in this article is to revisit the approach of Gillan *et al.* (14) using the Kool two-step task (7), which allows us to measure the effect of incentives on model-based control across a range of severity on several psychiatric constructs.

Principles of Metacontrol

On the basis of behavioral data from the Kool two-step task, we have argued that arbitration between model-based and model-free control is implemented by a cost-benefit analysis [see also Boureau *et al.* (20)]. According to this view, a metacontroller approximates the relative costs and benefits of using each controller and chooses one based on the optimal cost-benefit ratio (Figure 1A). In the Daw two-step task, model-based control confers no benefit, and hence the metacontroller will prefer the cognitively cheaper model-free controller. In the Kool two-step task, by contrast, model-based control does confer a benefit, so it will tend to be preferred by the metacontroller. This explains why people are overall more model based on the Kool two-step task than on the Daw two-step task, and why incentives increase reliance on model-based control only on the Kool two-step task.

We note that this is not the only way to understand the tension between habits and planning; for example, theories based on the free energy principle frame this tension in

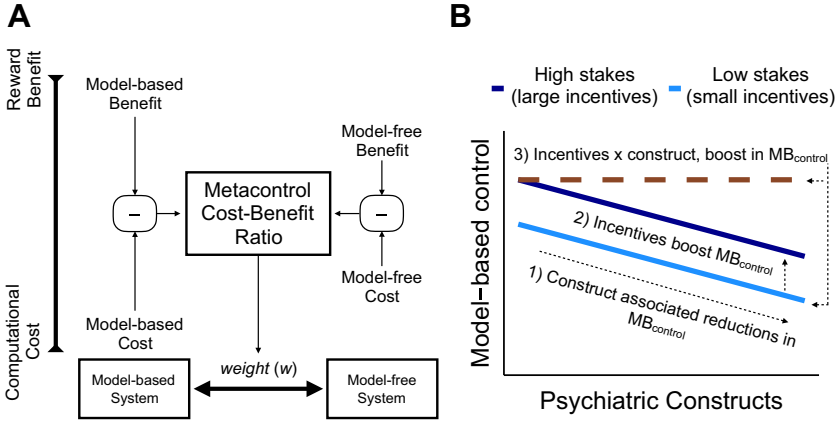

Figure 1. **(A)** Schematic showing how a metacontroller compares the computational cost against the reward benefit for model-based and model-free controllers. This produces a weighting parameter (w) for each individual that measures the probability of selecting model-based control ($MB_{control}$). **(B)** Schematic showing how $MB_{control}$ hypothetically changes with the degree of the psychiatric construct and incentives (high vs. low stakes).

terms of model comparison (21). However, because the model-based/model-free distinction is currently the most prominent and well-characterized framework, we focus on it in this article.

We use this framework to interpret the results of our study, by investigating three patterns (schematized in Figure 1B): 1) the effect of psychiatric constructs on model-based control, 2) the effect of incentives on model-based control, and 3) the interaction between these two effects. We expect, based on the work of Gillan *et al.* (14), that many psychiatric constructs will be associated with lower model-based control. Furthermore, we expect, based on the work of Kool *et al.* (7), that incentives will increase model-based control overall. The critical question is whether incentives manifest as a fixed boost in model-based control regardless of symptom or trait severity, or whether severity modulates the boost. The answer to this question will have implications for the effectiveness of incentive-based interventions. If high symptom or trait severity is associated with low sensitivity to incentives, then it is unlikely that such interventions will be effective for clinically relevant psychiatric symptoms and traits. If, on the other hand, the incentive effect observed by

Table 1. Self-reported Demographics (*N* = 839)

	%	*n*
Clinical Characteristic		
Endorsed specific diagnosis	36.83	309
2 diagnoses	13.59	114
≥3 diagnoses	6.80	57
Past Treatment		
Any treatment	30.99	260
Partial, inpatient, residential	8.46	72
Current Treatment		
Any treatment	19.79	166
Partial, inpatient, residential	4.29	36
Psychiatric medication	9.77	81

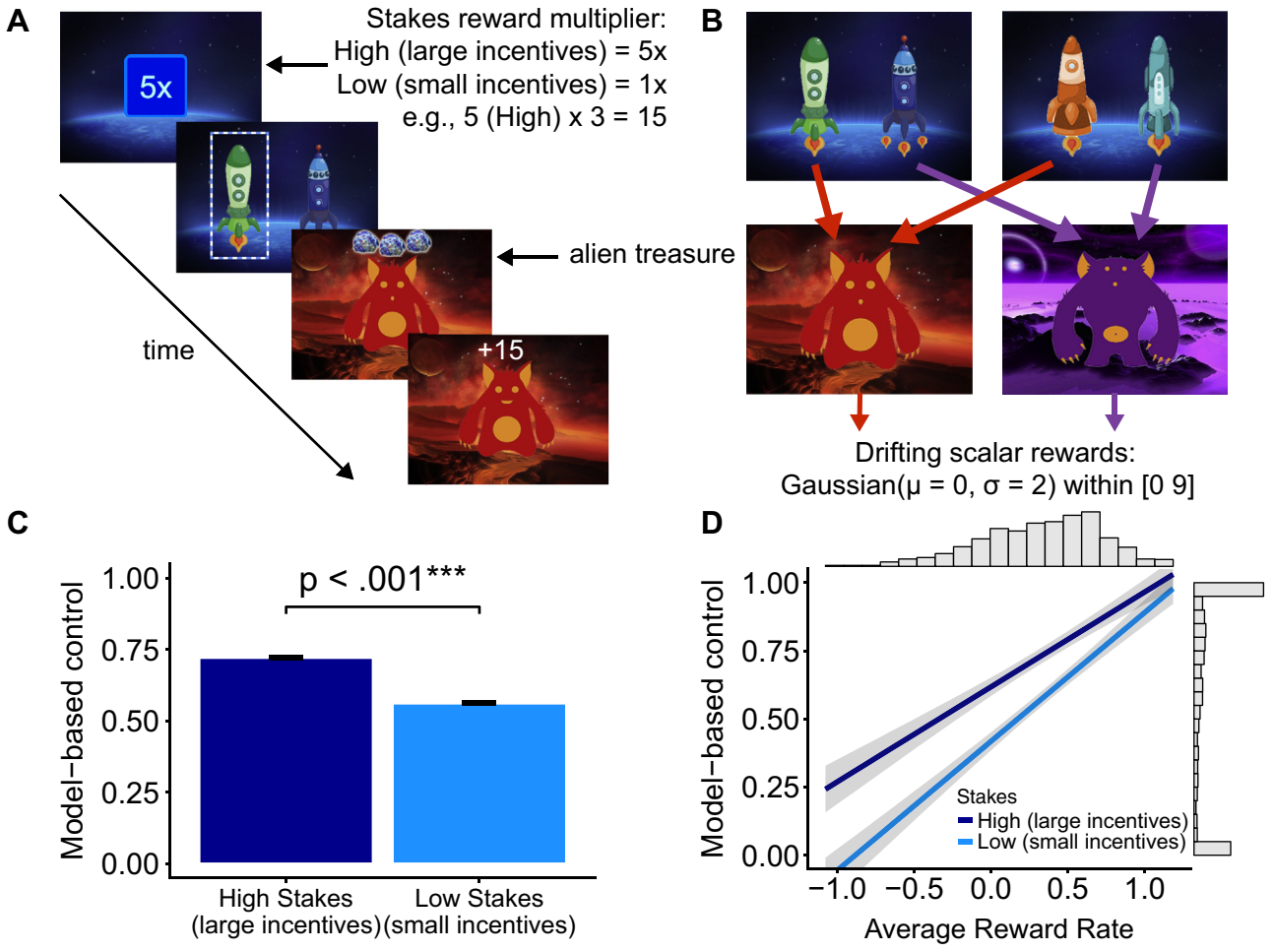

Figure 2. **(A)** On each trial, participants were first informed about the stakes, and then made a choice between two rockets. They then deterministically transitioned to a red or purple planet and received a reward. **(B)** State transition structure of the task. Rewards changed gradually over trials according to a Gaussian random walk. **(C)** Model-based control, measured by fitting a computational model to choice behavior, was significantly higher on high stakes trials compared with low stakes trials. **(D)** Greater model-based control was associated with higher average reward rate across stakes. Shading denotes credible intervals around β_{mean}.

Kool *et al.* (7) is independent of symptom or trait severity, then such interventions may hold promise.

METHODS AND MATERIALS

Participants

We recruited 941 participants using Amazon Mechanical Turk (Amazon, Seattle, WA). Participants gave informed consent and the study was approved by the Harvard Committee on the Use of Human Subjects.

Participants had to reside in the United States and have a 90% approval rating, and 100 participants completed Human Intelligence Tasks. Participants completed a computer-adaptive IQ test, the novel two-step paradigm, 19 clinical scales, a clinical demographic questionnaire (Table 1), and an additional cognitive task. Compensation was $20 with a performance bonus ($0.48 to $1.11). Following the exclusion criteria (Supplement), the final sample included 839 participants (48.75% women and 51.25% men) ranging from 18 to 73 years of age (mean age 34.95 ± 10.1 years) and with a mean IQ of 99.1 ± 9.71.

Self-report Measures

Participants completed several self-report measures (descriptions in Supplemental Table S1; mean [SD] data in Supplemental Table S2) including the Apathy Evaluation Scale (22), trait portion of the State-Trait Anxiety Inventory (23), Alcohol Use Disorders Identification Test (24), Zung Self-Rating Depression Scale (25), short schizotypy scale (26), Obsessive-Compulsive Inventory–Revised (27), Social Anxiety Scale (28), Eating Attitudes Test (29), Intolerance of Uncertainty Scale (30), Anxiety Sensitivity Index-3 (31), Ruminative Response Scale (32), Difficulties in Emotion Regulation Scale (33), Distress Tolerance Scale (34), Barratt Impulsiveness Scale-11 (35), and UPPS-P Impulsivity Scale (36–38) (the UPPS-P comprises positive and negative urgency, sensation seeking, lack of premeditation, and lack of perseverance). We summarized nine of these self-report measures (Supplemental Table S1) with three latent constructs (anxious depression, compulsive behavior and intrusive thought, and social withdrawal) that were generated using the factor loadings from Gillan *et al.* (14).

Sequential Decision Task

Participants performed 200 trials of the two-step task developed by Kool *et al.* (7), which allowed us to measure adaptive increases in model-based control by comparing the degree of model-based control across high and low stakes (Figure 2A, B). In this task, the participant randomly starts in one of two first-stage states, choosing one of two rocket ships (which are randomly mapped to response keys). Conditional on this choice, the participant deterministically transitions to the purple or red alien planet. On this alien planet, the participant then receives a reward indicated by alien treasure. The amount of reward obtained at each planet changes randomly and gradually over the task, independently for each planet (a Gaussian random walk between 0 and +9) (additional details in the Supplement).

If the participant fails to make a response, no reward is delivered and the task proceeds to the next trial. Each trial was randomly assigned to one of two conditions. In the high-stakes condition, the participant received five times the alien treasure reward (5× multiplier) (Figure 2A). In the low-stakes condition, the participant received the displayed alien treasure reward.

Analyses

Computational Model. We estimated model-based control for high stakes (large incentives) and low stakes (small incentives) with a dual-system reinforcement learning model (18,39). In this framework, model-free control learns the

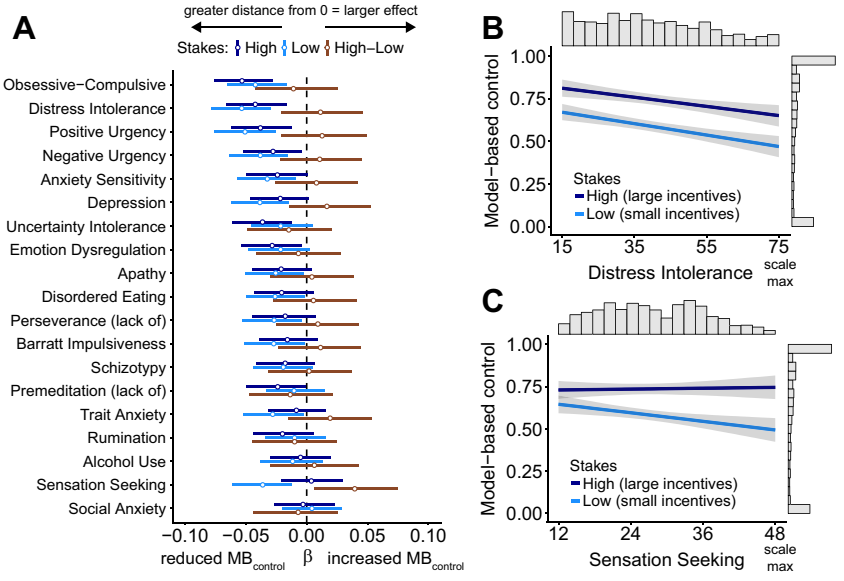

Figure 3. **(A)** Coefficient plot with credible intervals containing 95% of the posterior probability density around the mean, organized according to descending effect size. The coefficient value represents the estimated slope of the line relating symptom severity to model-based control ($MB_{control}$). Coefficients were estimated separately for high and low stakes; the high–low intervals show the relative effect. Distance from zero indicates a stronger relationship (e.g., stronger reductions in $MB_{control}$) and the credible interval indicates probable values of the self-report measure parameter estimate. Greater credible interval width indicates greater uncertainty about the parameter estimate. **(B)** Distress intolerance is associated with reduced model-based control, but high stakes boost model-based control by roughly the same amount regardless of the distress intolerance score. **(C)** Sensation seeking is associated with reduced model-based control in low stakes but not in high stakes.

value of actions through direct experience and is insensitive to sudden environmental changes. Alternatively, model-based control uses a model of the environment to instantaneously update which first-stage choice leads to each alien planet. The balance between model-free and model-based control is governed by a free parameter (w) in the computational model. We fit the parameter w for low stakes (small incentives) and high stakes (large incentives) separately (additional model fitting details in the Supplement).

Predictors of Model-Based Control (Bayesian Regression). Bayesian linear regression (40,41) was used to quantify the relationship among metacontrol, average reward rate, and self-report measures, while controlling for age, IQ, and gender. A classic general linear model results in a single beta value point estimate and confidence intervals around that beta value that represent the percentage of the time the interval would contain the population beta value if the study was conducted many times. In contrast, rather than providing a single point estimate, a Bayesian regression provides a posterior distribution of beta values given the data, quantifying the uncertainty about the betas. We summarized the posterior using a 95% highest posterior density credible interval around the mode. We also report the posterior probabilities that the beta value is greater or less zero.

We used the brms package (42) with the default prior $\sigma \sim$ *student-t*(3,0,10) and separately regressed each self-report measure and average reward rate onto model-based control (w) while controlling for age, IQ, and gender. We included a regressor expressing the interaction between the self-report measure and high or low stakes. Thus, we examined changes in model-based control across stakes as a function of each construct separately. This is consistent with Gillan *et al.* (14) and also retains the construct validity of these measures, as the constructs have been widely examined independently in prior research. We also ran a separate analysis entering all scales into a single regression, though it is possible the correlations across measures diminishes the construct validity of any individual measure. This is because it is unknown which portion of a single measures' variance accounts for the effects while controlling for other measures. In addition, we ran separate regressions adding the inverse temperature parameter as a confound.

Factor Scores. The three transdiagnostic factors were generated using the published factor loadings from Gillan *et al.* (14). The authors factor analyzed nine self-report scales reducing the data to three dimensions subsequently termed anxious depression, compulsive behavior and intrusive thought, and social withdrawal. Using the scales overlapping with Gillan *et al.* (14) (see Supplemental Table S1), we set their published factor loadings as independent variables in a regression and set the item level scores for our participants as the dependent variable, thereby generating factor scores for our participants. Using Bayesian regression, we entered factor scores for all three factors concurrently as independent variables and the interaction with stakes while controlling for age, IQ, and gender, with model-based control (w) as the dependent variable. Thus, we examined changes in model-based control (w) as a function of the interaction between stakes and the psychiatric factors. In a separate analysis, we also added inverse temperature as a covariate to mirror the self-report regressions.

RESULTS

Stakes, Reward Rate, and Model-Based Control

Consistent with the findings of Kool *et al.* (7), a paired-sample t test indicated participants engaged in significantly more model-based control (w) in the high-stakes condition compared with the low-stakes condition (Figure 2C) (t_{838} = 10.21, $p < .001$; Cohen's d = 0.35). Using Bayesian regression, the highest posterior density for high and low stakes indicated that the greater average reward rate predicted greater model-based control with high certainty (narrow credible intervals) (low-stakes condition [β_{mean} = .472, confidence interval = .417–.530], high-stakes condition [β_{mean} = .348, confidence interval = .296–.407]) (Figure 2D). These replicate the key findings from Kool *et al.* (7,19).

Self-report Measures and Model-Based Control

Figure 3A displays the mean beta value and credible interval for the association between self-report measures and model-based control. For the majority of constructs measured, Bayesian regressions revealed high certainty of the presence

Table 2. Posterior Probabilities of Reductions in Model-Based Control

	Construct		
	High Stakes Negative	Low Stakes Negative	High–Low Stakes Negative
Obsessive-Compulsive	100[a]	100[a]	73
Distress Intolerance	100[a]	100[a]	26
Positive Urgency	100[a]	100[a]	24
Negative Urgency	99[a]	100[a]	27
Anxiety Sensitivity	97[a]	99[a]	32
Depression	96[a]	100[a]	17
Uncertainty Intolerance	100[a]	95[a]	79
Emotion Dysregulation	99[a]	95[a]	65
Apathy	95[a]	98[a]	40
Disordered Eating	95	98[a]	38
Perseverance (Lack of)	91	98[a]	29
Barratt Impulsiveness	90	99[a]	27
Schizotypy	92	93	46
Premeditation (Lack of)	97[a]	80	79
Trait Anxiety	75	99[a]	14
Rumination	94	78	72
Alcohol Use	65	82	36
Sensation Seeking	38	100[a]	1[a]
Social Anxiety	59	37	66

The data are the probability (%) that each self-report measure predicts decreases in model-based control.

[a]At least 95% of the posterior probability density over β (the coefficient relating symptom severity to model-based control) is below zero (or above zero), indicating a high degree of certainty that symptom or trait severity is associated with reduced model-based control.

Table 3. Posterior Probabilities of Reductions in Model-Based Control

	Factor		
	High Stakes Negative	Low Stakes Negative	High–Low Stakes Negative
Anxious Depression	42	99[a]	4[a]
Compulsive-Intrusive	100[a]	100[a]	60
Social Withdrawal	7	0[a]	88

The data are the probability (%) that each transdiagnostic factor predicts decreases in model-based control. Psychiatric factors were generated using factor loadings from Gillan *et al.* (14).

[a]At least 95% of the posterior probability density over β (the coefficient relating transdiagnostic factor score to model-based control) is below zero (or above zero), indicating a high degree of certainty that the factor severity is associated with reduced model-based control.

of relationships with model-based control (*w*) (Table 2). For example, the posterior placed nearly all its mass on the beta value for distress intolerance (Figure 3B) being less than zero. This means that we can have high certainty that the relationship between distress intolerance and model-based control (*w*) is negative.

Of note, sensation seeking (Figure 3C) was the only measure showing an interaction with stakes, such that the large negative relationship between sensation seeking and model-based control (100% of distribution is highly negative) that was present during low stakes was not present during high stakes, with the posterior distribution roughly centered around zero (38% negative, 62% positive). Corresponding to this difference, 99% of the posterior distribution was positive for the high–low stakes interaction. Scatter plots showing original data points are in Supplemental Figure S1.

Additionally, the results of entering all measures into a single regression are in Supplemental Figure S4, but should be interpreted with caution because the correlations between measures (i.e., correlation heatmap) (Supplemental Figure S3) diminish their construct validity. Including inverse temperature largely reiterated the same pattern of results, though slightly weaker (Supplemental Figure S5). Nevertheless, inverse temperature and model-based control are highly correlated ($r_{1676} = -.4478$, $p < .001$) and suffer from nonidentifiability in the computational model. Therefore, including both builds unnecessary redundancy into the regression model.

Factors and Model-Based Control

Table 3 and Figure 4A show the mean beta values and credible intervals for the three transdiagnostic factors (i.e., compulsive behavior and intrusive thought, anxious depression, and social withdrawal) in predicting model-based control across stakes. The anxious-depression factor (Figure 4B) showed a similar pattern of results to sensation seeking, with 99% of the posterior probability for low stakes below zero, indicating a strong anticorrelation between anxious depression and model-based control; this effect was eliminated for high stakes, with the posterior probability closely centered around zero (42% negative, 58% positive). For the high- and low-stakes interaction, 96% of the posterior was positive. The compulsive behavior and intrusive thought factor (Figure 4C) posterior distribution was nearly 100% negative across stakes. Moreover, the mean beta value and credible interval demonstrated the largest effect in predicting reductions in model-based control relative to the other two factors. Using the 95% cutoff, the social withdrawal factor (Figure 4D) predicted increases in model-based control for low stakes, but this effect was slightly weakened for high stakes (93%). In a Bayesian framework, the high posterior probability for beta values below zero (e.g., compulsive behavior and intrusive thought factor) and large probabilities for beta values above zero (i.e., stakes interaction in anxious depression) provides relatively high

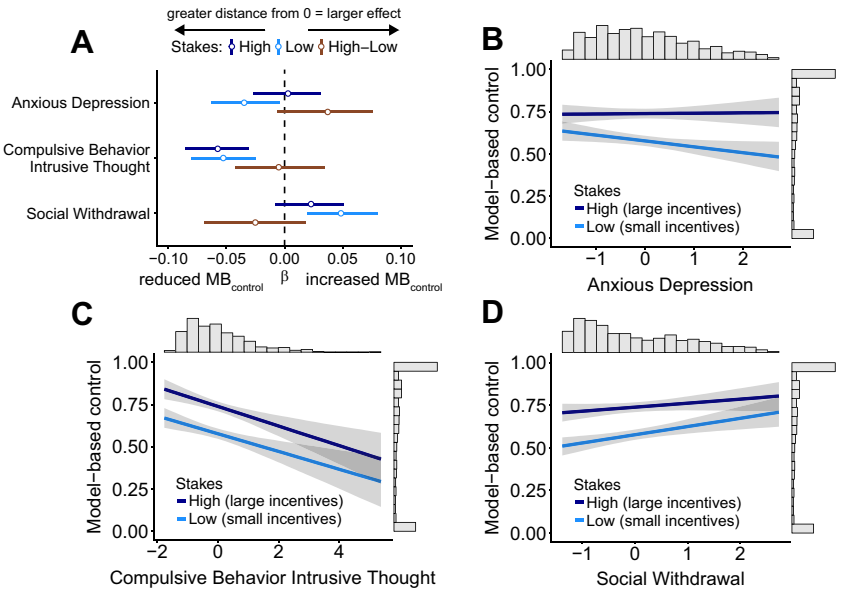

Figure 4. **(A)** Coefficient plot with credible intervals containing 95% of the posterior probability density around the mean for psychiatric factors in a single regression. Distance from zero indicates a stronger relationship (e.g., stronger reductions in model-based control [$MB_{control}$]) and the credible interval indicates probable values of the self-report measure parameter estimate. Greater credible interval width indicates greater uncertainty about the parameter estimate. **(B)** The anxious-depression factor is associated with reduced $MB_{control}$ for low stakes but not for high stakes. **(C)** The compulsive behavior and intrusive thought factor is associated with reduced model-based control, but high stakes boost model-based control regardless of the score. **(D)** The social withdrawal factor is associated with increased model-based control, further boosted by high stakes. Lines in panels **(B–D)** show regression lines with credible intervals.

certainty about the presence of three different relationships between the factors and model-based control. Scatter plots with original data points are displayed in Supplemental Figure S2. After adding inverse temperature to the model, the results remained consistent, though they were slightly weaker (Supplemental Figure S6).

DISCUSSION

The primary goal of the current study was to ask the question, Do people higher in symptom or trait severity boost model-based control in response to incentives? We found that 1) incentives boost model-based control across a range of severity on several psychiatric constructs; 2) sensation seeking and the anxious-depression factor showed a larger incentive effect, whereby higher severity was associated with greater boosts in model-based control; and 3) most constructs were associated with model-based deficits, but there were some exceptions, including social anxiety and alcohol use, which showed no relationship with model-based control, and the social withdrawal factor, which was related to increased model-based control.

In contrast to prior research on model-based control deficits in psychopathology, which has conceptualized these deficits as fixed individual traits (9–14), our metacontrol framework conceptualizes them as dynamic and adaptive. First, the results of the current study suggest that model-based control can be flexibly deployed depending on incentives in the environment, and increased clinical symptoms and traits do not diminish this flexibility. This has important clinical implications because interventions that use incentives have shown a range of positive psychiatric and health-related outcomes. For example, contingency management is an incentive-based intervention, whereby individuals diagnosed with substance use disorders are provided incentives in exchange for evidence of behavioral change (43). Moreover, incentives have been widely efficacious in fostering health-related goals such as smoking cessation, increased physical activity, healthy eating, and reduced alcohol consumption (44). Incentives also consistently improve response inhibition in patient populations diagnosed with attention-deficit/hyperactivity disorder (45). Potentially one reason these incentives have been found to be efficacious is that incentives may enhance processes underlying model-based control. Future studies could extend the current results in diagnosed patient samples by providing incentives for increases in model-based control, and testing if these changes correspond to treatment outcomes.

Second, our results showed an incentive interaction with both sensation seeking and the anxious-depression factor, in which increasing scores on these measures predicted larger boosts in model-based control in response to incentives. Although these two constructs show the same pattern of results, it may be due to different mechanisms. Sensation seeking is putatively characterized by increased appetitive drive and decreased sensitivity to punishment (46). One possibility is that the high-stakes condition may promote increased model-based control through hypersensitivity to reward, and this is consistent with theoretical models (46). For the anxious-depression factor, although it is unclear to us why this factor was related to larger boosts in model-based control, it seems unlikely that it is related to hypersensitivity to reward.

Third, the self-report measures generally showed a negative relationship with model-based control, with the exception of social anxiety and alcohol use, which were unrelated to model-based control, and sensation seeking, which showed an interaction with incentives. Some of the self-report measures had strong correlations (27% of the correlations were $r > .5$ but only 3% were $r > .7$) (see Supplemental Figure S3 for a correlation matrix). Thus, the negative relationship to model-based control across most constructs may be due to shared variance of a common illness factor (47,48). However, this is unlikely to be the sole explanation for our findings, because the majority of the correlations between the constructs were weak or moderate and, furthermore, the psychiatric factors show three separate relationships to model-based control (i.e., compulsive behavior shows a negative relationship, social withdrawal shows a positive relationship, and anxious depression shows an interaction with incentives).

Limitations

There are several limitations to the current study. First, our analyses were unable to tease apart whether deficits arise from impairment of the model-based controller itself, a dysfunctional metacontroller, or misrepresentation of the costs and benefits of choices. Future experimental and modeling work will be needed to tease apart these possibilities. Second, some constructs were sparsely sampled at the higher ranges, and given the current sample, we cannot make claims about whether incentives lead to similar boosts in model-based control among those at the highest end of severity (i.e., formally diagnosed and severe psychiatric patients). However, many of the participants here reported diagnoses and intensive psychiatric treatment (i.e., inpatient or residential care) (Table 1). Nevertheless, our data show that the incentive effect is largely uniform across the observed range for most constructs (Supplemental Figures S1 and S2). Third, based on our questions of interest, we did not test whether certain self-report measures are associated with model-based control, while simultaneously accounting for the effect of the other self-report measures, though we did take this approach with the factors. Future research could examine which constructs have unique relationships with the effects of incentives on model-based control.

Fourth, model-based control depends on a domain-general cognitive mechanisms, such as working memory (6,49). If a particular symptom is associated with a deficit in working memory, then this will manifest as reduced model-based control, but it will also manifest as a deficit in many other tasks. Future studies would have to measure (in the same individual) multiple tasks with overlapping cognitive demands to understand whether this is a major confound when examining model-based control and clinical constructs.

Last, one might be concerned that the task used here is “degenerate” as a sequential decision task, as there is only a single choice that results in deterministic transitions. Understanding whether our results generalize to the more complex sequential decision tasks characteristic of real-world environments remains an important unanswered question.

Conclusions

In this study, we found evidence for a cost-benefit metacontrol process that boosts model-based control under high incentive conditions regardless of symptom or trait severity for many psychiatric constructs. Exceptions included sensation seeking and the anxious-depression factor, for which higher construct severity was associated with larger boosts in model-based control for high-incentive conditions relative to low-incentive conditions. Alcohol use and social anxiety showed no relationship, while social withdrawal showed a positive relationship. The psychiatric factors revealed three separate relationships with model-based control, with the anxious-depression factor showing an interaction with incentives, thereby demonstrating the advantage of using psychiatric dimensions to detect specific relationships between self-report measures and computational constructs. Future work may seek to test the efficacy of using incentives to boost model-based control in patient populations and the relationship between these effects and treatment outcomes.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was partially supported by a training grant from the National Institutes of Health (NIH) Blueprint for Neuroscience Research Grant No. T90-DA022759/R90DA023427 (to EHP) and NIH Grant No. CRCNS R01-1207833 (to SJG). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

We would like to thank Catherine Hartley for sharing her stimuli, Michal Kosinksi for use of the IQ test, Drs. Bruce Rosen and Maria Mody for their work with the Advanced Multimodal Neuroimaging Training Program, Ista Zahn for his helpful feedback during the revision process, and the members of the Computational Cognitive Neuroscience Laboratory for their support and feedback.

The authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Department of Psychology (EHP, WK, AJM, SJG) and Center for Brain Science (EHP, SJG), Harvard University, Cambridge, Massachusetts.

Address correspondence to Edward H. Patzelt, M.A., Harvard University, 52 Oxford St, Northwest Building, 295.07, Cambridge, MA 02138; E-mail: patzelt@g.harvard.edu.

Received Oct 4, 2017; revised May 8, 2018; accepted Jun 19, 2018.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.biopsych.2018.06.018.

REFERENCES

1. Everitt BJ, Robbins TW (2005): Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. Nat Neurosci 8:1481–1489.
2. Daw ND, Niv Y, Dayan P (2005): Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704–1711.
3. Dolan RJ, Dayan P (2013): Goals and habits in the brain. Neuron 80:312–325.
4. Kool W, Cushman FA, Gershman SJ (2018): Competition and cooperation between multiple reinforcement learning systems. In: Morris RW, Bornstein AM, Shenhav A, editors. Understanding Goal-Directed Decision Making: Computations and Circuits. New York: Elsevier.
5. Gershman SJ, Markman AB, Otto AR (2014): Retrospective revaluation in sequential decision making: A tale of two systems. J Exp Psychol Gen 143:182–194.
6. Otto AR, Gershman SJ, Markman AB, Daw ND (2013): The curse of planning: Dissecting multiple reinforcement-learning systems by taxing the central executive. Psychol Sci 24:751–761.
7. Kool W, Gershman SJ, Cushman F (2017): Cost-benefit arbitration between multiple reinforcement learning systems. Psychol Sci 28:1321–1333.
8. Wan Lee S, Shimojo S, O'Doherty JP (2014): Neural computations underlying arbitration between model-based and model-free learning. Neuron 81:687–699.
9. Culbreth AJ, Westbrook A, Daw ND, Botvinick M, Barch DM (2016): Reduced model-based decision-making in schizophrenia. J Abnorm Psychol 125:777–787.
10. Voon V, Derbyshire K, Rück C, Irvine MA, Worbe Y, Enander J, *et al.* (2015): Disorders of compulsivity: A common bias towards learning habits. Mol Psychiatry 20:345–352.
11. Sebold M, Deserno L, Nebe S, Schad DJ, Garbusow M, Hägele C, *et al.* (2014): Model-based and model-free decisions in alcohol dependence. Neuropsychobiology 70:122–131.
12. Sebold M, Nebe S, Garbusow M, Guggenmos M, Schad DJ, Beck A, *et al.* (2017): When habits are dangerous: Alcohol expectancies and habitual decision making predict relapse in alcohol dependence. Biol Psychiatry 82:847–856.
13. Nebe S, Kroemer NB, Schad DJ, Bernhardt N, Sebold M, Müller DK, *et al.* (2017): No association of goal-directed and habitual control with alcohol consumption in young adults. Addict Biol 23:379–393.
14. Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND (2016): Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. Elife 5:e11305.
15. Cuthbert BN, Insel TR (2013): Toward the future of psychiatric diagnosis: The seven pillars of RDoC. BMC Med 11:126.
16. Insel T, Cuthbert B, Garvie M, Heinssen R, Pine DS, Quinn K, *et al.* (2010): Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. Am J Psychiatry 167:748–751.
17. Clark LA, Cuthbert B, Lewis-Fernández R, Narrow WE, Reed GM (2017): Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). Psychol Sci Public Interes 18: 72–145.
18. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011): Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204–1215.
19. Kool W, Cushman FA, Gershman SJ (2016): When does model-based control pay off? PLoS Comput Biol 12:e1005090.
20. Boureau YL, Sokol-Hessner P, Daw ND (2015): Deciding how to decide: Self-control and meta-decision making. Trends Cogn Sci 19:700–710.
21. FitzGerald THB, Dolan RJ, Friston KJ (2014): Model averaging, optimal inference, and habit formation. Front Hum Neurosci 8:457.
22. Marin R, Biedrzycki R, Firinciogullari S (1991): Reliability and validity of the apathy evaluation scale. Psychiatry Res 38:143–162.
23. Spielberger CD, Gorsuch RL, Lushene PR, Vagg PR, Jacobs AG (1983): Manual for the State–Trait Anxiety Inventory. Sunnyvale, CA: Consulting Psychologists Press.
24. Saunders JB, Aasland OG, Babor TF, De La Fuente JR, Grant M (1993): Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption. II. Addiction 88:791–804.
25. Zung W, Durham N (1965): A self-rating depression scale. Arch Gen Psychiatry 12:63–70.
26. Mason O, Linney Y, Claridge G (2005): Short scales for measuring schizotypy. Schizophr Res 78:293–296.
27. Foa EB, Huppert JD, Leiberg S, Langner R, Kichic R, Hajcak G, Salkovskis PM (2002): The obsessive-compulsive inventory: Development and validation of a short version. Psychol Assess 14:485–496.
28. Liebowitz MR (1987): Social phobia. Mod Probl Pharmacopsychiatry 22:141–173.

29. Garner D, Olmsted MP, Bohr Y, Garfinkel P (1982): The eating attitudes test: Psychometric features and clinical correlates. Psychol Med 12:871–878.
30. Carleton RN, Norton MAPJ, Asmundson JG (2007): Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. J Anxiety Disord 21:105–117.
31. Taylor S, Zvolensky MJ, Cox BJ, Deacon B, Heimberg RG, Ledley DR, *et al.* (2007): Robust dimensions of anxiety sensitivity: Development and initial validation of the Anxiety Sensitivity Index-3. Psychol Assess 19:176–188.
32. Nolen-Hoeksema S, Morrow J (1991): A prospective study of depression and posttraumatic stress symptoms after a natural disaster: The 1989 Loma Prieta earthquake. J Pers Soc Psychol 61:115–121.
33. Gratz KL, Roemer L (2004): Multidimensional assessment of emotion regulation and dysregulation. J Psychopathol Behav Assess 26: 41–54.
34. Simons JS, Gaher RM (2005): The distress tolerance scale: Development and validation of a self-report measure. Motiv Emot 29: 83–102.
35. Patton JH, Stanford MS, Barratt ES (1995): Factor structure of the Barratt impulsiveness scale. J Clin Psychol 51:768–774.
36. Lynam DR, Smith GT, Cyders MA, Fischer S, Whiteside SP (2007): The UPPS-P Questionnaire Measure of Five Dispositions to Rash Action. Unpublished Technical Report. West Lafayette, IN: Purdue University.
37. Cyders MA, Smith GT, Spillane NS, Fischer S, Annus AM, Peterson C (2007): Integration of impulsivity and positive mood to predict risky behavior: Development and validation of a measure of positive urgency. Psychol Assess 19:107–118.
38. Whiteside SP, Lynam DR (2001): The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. Pers Individ Dif 30:669–689.
39. Sutton RS, Barto AG (1998): Reinforcement Learning: An Introduction. Cambridge, MA: MIT Press.
40. Baldwin SA, Larson MJ (2016): An introduction to using Bayesian linear regression with clinical data. Behav Res Ther 98:58–75.
41. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013): Bayesian Data Analysis, 3rd ed (Chapman Hall/CRC Texts Statistical Science). Boca Raton; FL: CRC Press.
42. Burkner PC (2013): brms: An R package for Bayesian multilevel models using Stan. J Stat Softw 80:1–28.
43. Davis DR, Kurti AN, Skelly JM, Redner R, White TJ, Higgins ST (2016): A review of the literature on contingency management in the treatment of substance use disorders, 2009–2014. Prev Med 92: 36–46.
44. Mantzari E, Vogt F, Shemilt I, Wei Y, Higgins JPT, Marteau TM (2015): Personal financial incentives for changing habitual health-related behaviors: A systematic review and meta-analysis. Prev Med 75:75–85.
45. Ma I, van Duijvenvoorde A, Scheres A (2016): The interaction between reinforcement and inhibitory control in ADHD: A review and research guidelines. Clin Psychol Rev 44:94–111.
46. Rochat L, Billieux J, Gagnon J, Van der Linden M (2018): A multifactorial and integrative approach to impulsivity in neuropsychology: Insights from the UPPS model of impulsivity. J Clin Exp Neuropsychol 40:45–61.
47. Caspi A, Moffitt TE (2018): All for one and one for all: Mental disorders in one dimension [published online ahead of print Apr 6]. Am J Psychiatry.
48. Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, *et al.* (2014): The p Factor: One general psychopathology factor in the structure of psychiatric disorders? Clin Psychol Sci 2:119–137.
49. Otto AR, Raio CM, Chiang A, Phelps EA, Daw ND (2013): Working-memory capacity protects model-based learning from stress. Proc Natl Acad Sci U S A 110:20941–20946.