Working-memory capacity protects model-based learning from stress

A. Ross Otto[a,1], Candace M. Raio[b], Alice Chiang[b], Elizabeth A. Phelps[a,b,c], and Nathaniel D. Daw[a,b]

[a]Center for Neural Science and [b]Department of Psychology, New York University, New York, NY 10003; and [c]Nathan Kline Institute, Orangeburg, NY 10962

Edited by Bruce S. McEwen, The Rockefeller University, New York, NY, and approved November 11, 2013 (received for review July 11, 2013)

Accounts of decision-making have long posited the operation of separate, competing valuation systems in the control of choice behavior. Recent theoretical and experimental advances suggest that this classic distinction between habitual and goal-directed (or more generally, automatic and controlled) choice may arise from two computational strategies for reinforcement learning, called model-free and model-based learning. Popular neurocomputational accounts of reward processing emphasize the involvement of the dopaminergic system in model-free learning and prefrontal, central executive–dependent control systems in model-based choice. Here we hypothesized that the hypothalamic-pituitary-adrenal (HPA) axis stress response—believed to have detrimental effects on prefrontal cortex function—should selectively attenuate model-based contributions to behavior. To test this, we paired an acute stressor with a sequential decision-making task that affords distinguishing the relative contributions of the two learning strategies. We assessed baseline working-memory (WM) capacity and used salivary cortisol levels to measure HPA axis stress response. We found that stress response attenuates the contribution of model-based, but not model-free, contributions to behavior. Moreover, stress-induced behavioral changes were modulated by individual WM capacity, such that low-WM-capacity individuals were more susceptible to detrimental stress effects than high-WM-capacity individuals. These results enrich existing accounts of the interplay between acute stress, working memory, and prefrontal function and suggest that executive function may be protective against the deleterious effects of acute stress.

A number of accounts of human and animal decision-making posit the coexistence of separate valuation systems that control choice (1–4), which, broadly speaking, represent automatic or habitual vs. deliberative or controlled modes. The circumstances under which one system may dominate over the other and thereby exert control over behavior has been a question of interest in both neuroscience and psychology, in part because of the implications of such differential control for disorders of compulsion such as drug abuse (5, 6). Acute stress may afford unique leverage in isolating the properties of these systems, because it is believed to prompt a shift from more cognitive or deliberative processes to more automatic processes presumed to be underpinned by phylogenetically older brain structures (7).

Accordingly, a spate of recent work suggests that acute stress—indexed by changes in levels of cortisol, a neuroendocrine marker of stress response—engenders reliance on putative habitual and/or automatic processes in human decision-making (8–13), consistent with the assumption that the physiological stress response impairs central executive functions subserving more deliberative choice. However, distinguishing such processes is both experimentally and theoretically fraught, because in dual process theories, which system controls a particular behavior is typically ambiguous, and can only be recognized by characteristics (such as reaction times or conscious access) associated in different theories with either sort of control, and often only in the comparison between different tasks that promote either mode. Here we leverage a more operational version of this distinction based on reinforcement learning (RL) theory (1), which proposes that deliberative and automatic modes of decision-making arise from two distinct computationally precise and neurobiologically grounded learning strategies for evaluating actions from previous experiences. This approach allows us to characterize more precisely and within a single task the impact of physiological stress response upon trial-by-trial learning dynamics of either sort.

This RL framework (1) posits that choice behavior arises from a combination of two value learning systems that operate in parallel and whose fundamental difference is whether they rely on an "internal model" of task contingencies for evaluating choices. The model-based system is computationally sophisticated and learns a model of the environment to plan candidate courses of action prospectively. In contrast, the model-free system eschews this model and merely prescribes that previously rewarded actions are repeated, akin to the Law of Effect and to prominent theories in which dopaminergic prediction error responses drive learning about action preferences at target areas such as the striatum (14, 15). Because these hypothesized modes of choice are defined quantitatively as arising from different trial-by-trial learning rules, they make clear and divergent predictions about subjects' trial-by-trial adjustment of decision preferences in response to feedback, enabling the contributions of both approaches to be dissociated experimentally. In fact, many laboratory choice tasks cannot differentiate between the contributions of the learning strategies, because when each action is paired with a single reward, the two sorts of value learning reduce to the same learning rule. However, the strategies differ appreciably in sequentially structured choice tasks. Recent work, informed by this approach, reveals that under normal circumstances, human learning in such tasks exhibits contributions of both putative systems (16–18). The grounding of these theories in neurocomputational models (19) and work on animal learning (4) also provides a unique perspective on dual process architectures,

Significance

The physiological response evoked by short-lived stressful events, referred to as acute stress, impacts human decision-making. Past studies assume that stress causes people to fall back, from more cognitive or deliberative modes of choice, to more primitive or automatic modes of choice because stress impairs peoples' capacity to process information (working memory). We directly examined how acute stress affects choice in a laboratory decision-making task for which the working memory demands of the two forms of decision-making are well understood, finding that stress impaired use of sophisticated choice strategies that require working memory but did not affect use of simpler, more primitive strategies. Further, this impairment was exacerbated in individuals with smaller working memory capacity, which is related to general intelligence.

Author contributions: A.R.O., C.M.R., E.A.P., and N.D.D. designed research; A.R.O. and A.C. performed research; A.R.O. analyzed data; and A.R.O., C.M.R., E.A.P., and N.D.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1] To whom correspondence should be addressed. E-mail: rotto@nyu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1312011110/-/DCSupplemental.

complementary to a set of views whose roots lie more in human cognitive neuroscience.

In line with the considerable computational requirements of model-based evaluation (1, 20), and with evidence that this process relies on the prefrontal cortex (PFC, 4), recent work suggests that the model-based system imposes considerable demands on central executive resources. In particular, depletion of working-memory (WM) resources abolishes model-based contributions to learning behavior but spares model-free contributions (21). At the same time, a different line of work examining central executive function under acute stress reveals how neurophysiological stress response engenders WM capacity impairment (22, 23) and reduction of WM-related activity in the PFC as assessed by neuroimaging (24).

On the basis of these two lines of work, an intuitive prediction emerges: stress response—as it deleteriously impacts the PFC-dependent executive resources—should selectively reduce model-based learning, but simultaneously spare model-free learning. Closely supporting this prediction, previous investigations reveal that acute stress engenders reliance on habitual behaviors, at the expense of flexible, goal-directed responding. However, because the two forms of choice were differentiated by posttraining probe trials—testing flexible sensitivity to reinforcer devaluation (25) or to a conjunction of spatial cues (26)—it remains to be investigated how and whether stress affects either of the two sorts of trial-by-trial learning dynamics that have been hypothesized to give rise to the endpoint behaviors probed there (1).

A complimentary possibility is suggested by findings that acute stress can increase firing rates of dopaminergic neurons (27) and extracellular dopamine levels in the neural structures putatively underpinning model-free RL (28). We might thus expect, alternately or additionally, that stress would modulate or even strengthen model-free learning. There is indeed recent evidence for effects of stress on probabilistic reward learning (29, 30). However, the task used does not permit differentiating model-based from model-free contributions to learning.

Here we elucidate the impact of hypothalamic-pituitary-adrenal (HPA) axis stress response on the expression of model-based and model-free contributions to sequential choice behavior. In the RL task we use (16) model-based and model-free learning strategies—distinguished, respectively, by their utilization and ignorance of the full environment structure—that give rise to distinct and quantifiable behavioral signatures. Our results reveal how the physiological stress response attenuates the influence of model-based (but not model-free) learning, underlining the distinct and separable contributions of these theorized valuation systems.

Further, in line with the central-executive–dependent nature of the model-based system, we shed light on how individual differences in WM capacity (often taken as a general measure of executive function and fluid intelligence, 31), modulate the effect of physiological stress response on model-based choice. Specifically, we demonstrate that subjects with more executive resources to spare find themselves less susceptible to the behavioral changes brought about by stress response, elucidating the interplay between acute stress, executive function, and dual-system accounts of decision-making.

Results

Subjects performed 200 trials of a two-step RL task (Fig. 1) (22, 29), designed to dissociate model-free and model-based learning strategies. In each two-stage trial, subjects made an initial first-stage choice between two options (depicted as fractals), which probabilistically leads to one of two second-stage states (colored green or blue). In each of these subsequent states, subjects made another choice between two options, which were associated with different probabilities of monetary reward. Choosing the left action at the first stage usually leads to the green state (70% of the time, a common transition) but sometimes leads to the blue state (30% of the time, a rare transition). Because the reward probabilities associated with second-stage choices drift over time according to independent random walks, subjects need to make trial-by-trial adjustments to their choices at both stages to effectively maximize payoffs.

Model-based and model-free strategies make different predictions about how the history of rewards received at the second stage should influence first-stage choices, owing to the fact that the model-free approach evaluates actions retrospectively, by learning to repeat actions that tend to be rewarded, whereas model-based learning evaluates actions prospectively, in terms a learned model of their likely consequences. For example, consider a first-stage choice that results in a rare transition to a second-stage state, and the subsequent second-stage choice is rewarded. Under a pure model-free strategy, by virtue of the reinforcement principle [or the temporal difference (TD) algorithm (λ) for $\lambda > 0$], one would have an increased chance of repeating the same first-stage response because it ultimately resulted in reward. In contrast, a model-based strategy—using a model of the task's transition structure—predicts a decreased tendency to repeat the same first-stage action because the other first-stage action is the one that is more likely to lead to that rewarded second-stage state.

Accordingly, below we examine how stress alters the learning systems' contributions by examining trial-by-trial adjustments in choices as subjects receive feedback. First, by formalizing each system's learning (1) with a trial-by-trial mathematical model and fitting it to subjects' choices, we measure how stress response affects the relative expression of the two learning systems. Next, we probe how stress impacts more qualitative signatures of each system, by examining trial-by-trial staying or switching in

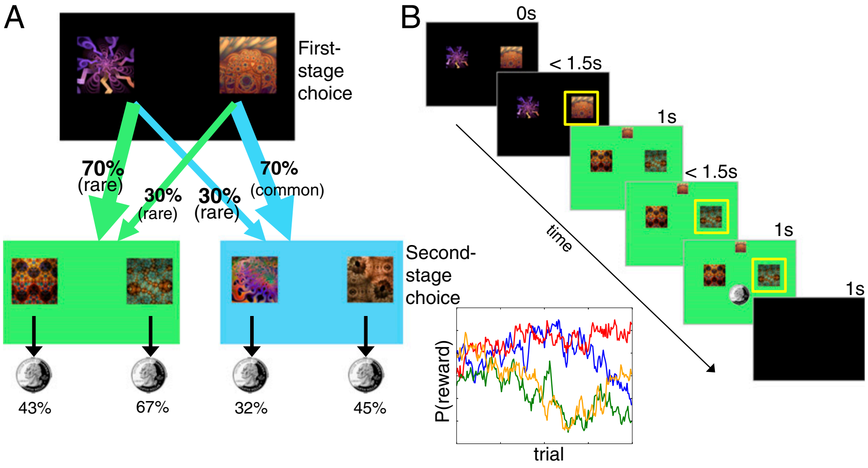

Fig. 1. State transition and reward structure in the two-step RL task. Each first-stage choice (black background) is predominantly associated with one or the other of the second-stage states (green and blue backgrounds) and leads there 70% of the time. These second-stage choices are probabilistically reinforced with money, whose reward probabilities change over the course of the experiment (see *Results* for a detailed explanation).

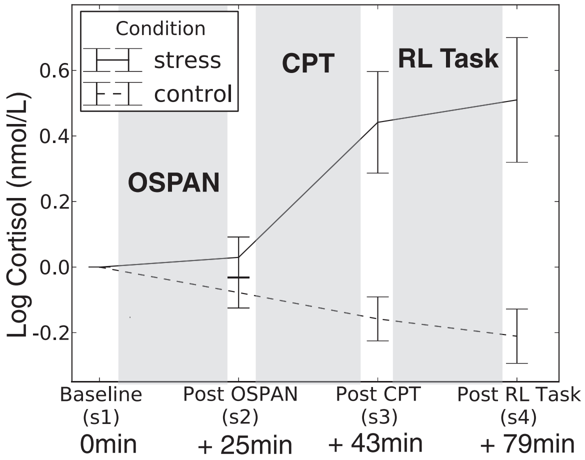

Fig. 2. Cortisol was significantly elevated among stress subjects, relative to controls, at the two time points following administration of the CPT (s3 and s4). Error bars denote SEM.

response to choice outcomes, for which the two accounts of learning predict different patterns.

Physiological and Subjective Response to Stress. We manipulated stress levels by having subjects undergo either the cold pressor test (CPT) task (32), an acute stress induction in which subjects submerged their arms in ice water for 3 min, or a control task using room temperature water. Baseline-subtracted cortisol concentrations over the four samples are plotted in Fig. 2 (raw data provided in Table S1). Critically, we found a significant interaction between condition (stress/control) and time of cortisol measurement ($F = 19.99$, $P < 0.0001$), indicating that the acute stressor induced a marked cortisol response. Moreover, within groups, cortisol concentrations did not change significantly between s3 and s4 ($P > 0.54$), suggesting cortisol concentrations remained steady throughout the RL task. Subjects in the stress condition reported that the CPT was significantly more unpleasant [mean (M) = 6.68, SD = 0.54] than control subjects (M = 2.19, SD = 0.38, $t = 6.95$, $P < 0.001$), indicating that the manipulation evoked a subjective stress response.

Stress Response and Model-Based Behavioral Contributions. We fit a dual-system RL model—a computational instantiation of the principles governing two hypothesized choice systems (16, 18)—to subjects' trial-by-trial choices (*SI Text*). This model consists of a model-free system that updates estimates of choice values using TD learning and a model-based system that learns a transition and reward model of the task and uses these to compute choice values on the fly. The model includes parameters controlling the influence of each system in determining choice and a learning rate controlling the decay timescale over which past rewards are considered in the systems' learning. We used hierarchical Bayesian model-fitting techniques (33) to estimate these parameters (Table S2).

Critically, how subjects adjust their trial-by-trial choice preferences in response to feedback reveals the extent to which they rely on either system. Table 1 reports the estimated parameters. Mirroring findings from previous work (16, 21, 34, 35) that both strategies influenced behavior, β_{MF} and β_{MB}, which quantify the weight given to model-free and model-based values in determining choice, respectively, were both significantly positive. We additionally estimated the extent to which each of these parameters changed with the stress response (quantified by cortisol delta; *Materials and Methods*). The parameter β_{MBcort}, which quantifies change in model-based contribution as a function of an individual subject's stress response, was significantly negative, indicating that model-based contributions decreased as stress response increased (Fig. 3*A*). We further hypothesized that, as model-free choice does not impose the same requirements on central executive resources as model-based choice, stress should not impact the contribution of a model-free strategy. Indeed, β_{MFcort} was not significantly different from zero, indicating that cortisol delta (i.e., stress response) did not alter model-free contributions (Fig. 3*B*).

Individual Differences in WM Capacity. We examined how individual WM capacity—operationalized by Operation Span (OSPAN; *Materials and Methods*)—modulates the effect of cortisol response on model-based choice. The parameter β_{MBcxo}, which quantifies the change in model-based contribution as a function of the interaction between OSPAN and stress response, was significantly positive. The positive relationship indicates that subjects with lower WM capacities were more susceptible to the effect of cortisol delta on model-based choice contribution, whereas subjects with higher WM capacities were effectively shielded from this effect.

This relationship is visualized in Fig. 4, by dividing subjects into low and high WM capacities according to a median split: among subjects low in WM capacity, cortisol delta reduced the expression of model-based choice (Fig. 4*A*), but among subjects high in WM capacity, cortisol response did not produce an appreciable impact on model-based contributions to behavior (Fig. 4*B*). Furthermore, as the predicted locus of the OSPAN shielding effect is model-based (but not model-free) contributions, we found that OSPAN did not interact significantly with the relationship between cortisol response and previous reward, the marker for model-free learning (i.e., β_{MFcxo} was not significantly different from zero; Table 1).

Logistic Regression Analysis. To more directly characterize the effect of stress and OSPAN on learning, we examined how the outcomes of each trial impact the next trial's choice, an approach taken in previous work (16, 21, 32). This restricted analysis

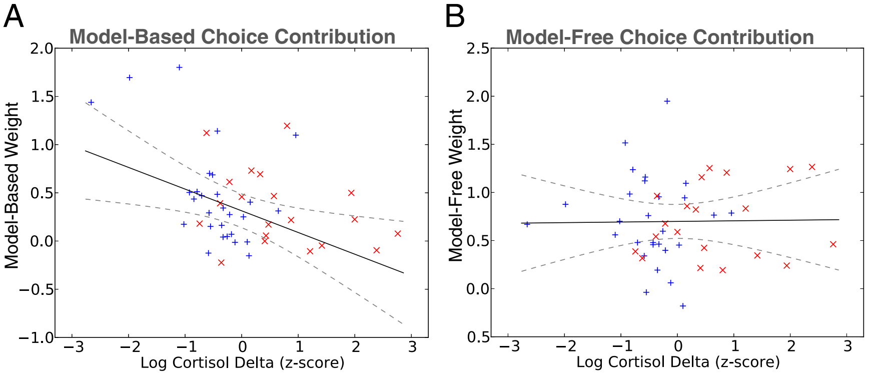

Fig. 3. Effect of stress on model-based vs. model-free value weights, as determined by the computational model. (*A*) Individual subjects' model-based value weights, plotted separately for subjects in the control and stress conditions. There was a significant negative effect of cortisol delta on expression of model-based learning, indicating cortisol change diminished its behavioral expression. (*B*) Model-free contribution to behavior. Note that there was no significant effect of cortisol change on expression of model-free choice, indicating that expression of model-free contribution is spared. Regression lines are computed from the population-level estimate of the log-linear effect of stress on model-based weight. Dashed gray lines indicate 2 SEs.

Table 1. Medians and 95% CI boundaries for the four parameters of interest, relating stress and OSPAN to model-based and model-free contributions

Parameter	Description	Median	Lower 95%	Upper 95%
β_{MB}	Model-based weight	0.313	0.149	0.492
β_{MF}	Model-free weight	0.693	0.533	0.88
β_{MBcort}	Cortisol effect on model-based weight	−0.2261	−0.4011	−0.0552
β_{MFcort}	Cortisol effect on model-free weight	0.0069	−0.1693	0.1870
β_{MBcxo}	Cortisol × OSPAN effect on model-based weight	0.4201	0.1341	0.7508
β_{MFcxo}	Cortisol × OSPAN effect on model-free weight	0.0211	−0.2291	0.2723

permits a more direct and qualitative examination of model-based and model-free contributions to trial-by-trial learning, because the two strategies make qualitatively distinct predictions about how the reward (rewarded vs. unrewarded) and transition type (common vs. rare) on the immediately preceding trial should influence first-stage choices (Fig. 5). A pure model-free strategy prescribes that the previous reward should influence whether a first-stage action is repeated, independent of which state (common or rare) it was received in. Thus, this algorithm predicts only a main effect of reward. In contrast, a model-based strategy predicts an interaction between the two factors because the effect of the previous reward on the first-stage choice depends on which state it was received. Note that, although both systems (in principle and empirically as estimated above) learn incrementally so that multiple preceding trials' outcomes influence each choice, these qualitative effects of the single most recent trial still hold.

The regression analysis confirmed the basic signatures of model-free and model-based strategies as described by the computational model, expressed as significant effects of both previous reward and the interaction between previous reward and transition type ($P < 0.001$; Table 2, Table S3, and Fig. S1). Moreover, the regression revealed that stress effectively attenuated the model-based learning, expressed as the negative interaction between cortisol response, previous reward, and transition type ($P < 0.01$), but not model-free learning, expressed as the simple effect of previous reward ($P > 0.5$).

We also specified a model examining how cortisol delta and OSPAN interacted with the same trial-by-trial variables in the above analysis (Table S4). Critically, OSPAN significantly interacted with the three-way interaction between cortisol response, previous reward, and previous transition type (the interaction signifying cortisol response's effect on model-based choice, $P < 0.01$). The positive coefficient indicates that subjects with lower WM capacities were more susceptible to stress' effects on model-based learning, corroborating the computational model fits (Fig. S2).

Discussion

Although a recent body of work has sought to understand the impact of stress on decision-making through a dual-systems framework (10, 36), in the absence of a clear computational framework for valuation, it is difficult to determine the locus of the stress-induced breakdown. Recent work (16, 18) suggests that sequential choice results from two distinct learning strategies for determining choice value from previous experience. Moreover, although dual-process accounts in psychology emphasize the role of WM capacity in determining reliance on and behavioral expression of the two systems (37, 38), the dependence of the two hypothesized modes of choice on central executive resources is not well understood (39). Leveraging a contemporary RL-based framework (1) in which the behavioral contributions of model-based and model-free strategies are separately identifiable and their differential demands on the central executive resources have been characterized (21), we reveal how neurophysiological stress response diminishes the contribution of a computationally expensive, model-based choice strategy but leaves intact the contribution of the more parsimonious model-free valuation system. This approach yields a rich picture of acute stress' impact on decision-making as it ties together lines of work examining stress response and PFC-dependent executive functions and dual-system theories of choice.

Perhaps more striking is that individual WM capacity—closely related to fluid intelligence and general cognitive ability (31)—appears to protect decision-makers from the deleterious effects of stress response. That is, we found that cortisol reactivity hampers the expression of model-based choice in low, but not high, WM capacity individuals. This result dovetails with notion of "cognitive reserve" in neuropsychology (40). On this view, individual differences in cognitive ability (often operationalized as IQ) allow some people to cope better than others with brain insult. It is conceivable, in the present study, that individuals with greater processing capacity (indexed here by OSPAN) were less burdened by the computational expense of model-based choice (20) and thus, found their choices less severely impacted by HPA axis response. Indeed, such individual differences could elucidate the considerable heterogeneity found in stress-induced changes to decision-making (36): individuals with

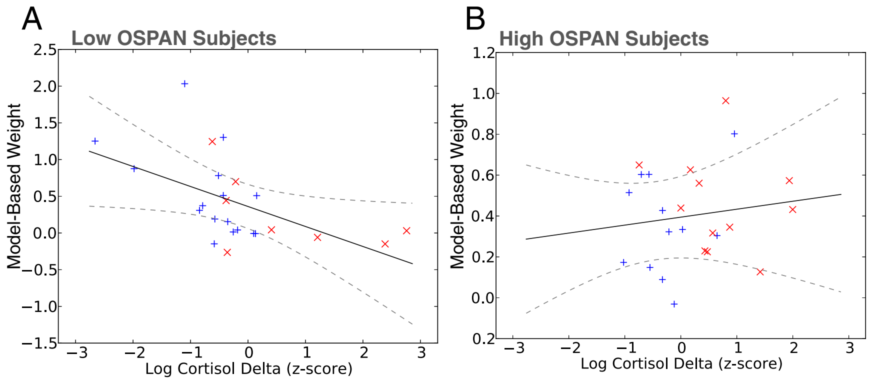

Fig. 4. Effect of stress on model-based learning as a function of individual WM capacity, as measured by OSPAN. Individual subjects' model-based value weights are plotted for low OSPAN subjects (*A*) and high OSPAN subjects (*B*). Cortisol response markedly dampened expression of model-based choice in the low OSPAN subgroup but not in the high OSPAN subgroup. Regression lines are computed from the population-level estimate of the log-linear effect of stress on model-based weight. Dashed gray lines indicate 2 SEs.

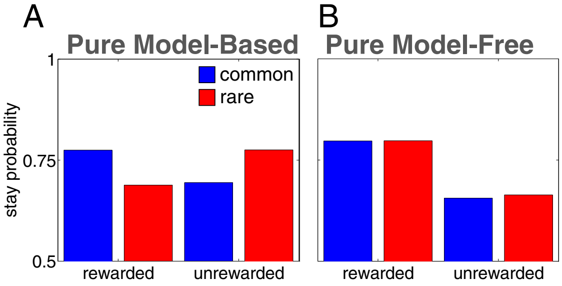

Fig. 5. (*A*) A model-based based choice strategy predicts that rewards after rare transitions should affect the value of the unchosen first-stage option, leading to a predicted interaction between the factors of reward and transition probability. (*B*) In contrast, a model-free strategy predicts that a first-stage choice resulting in reward is more likely to be repeated on the subsequent trial regardless of whether that reward occurred after a common or rare transition.

larger executive capacities could find their behavior less compromised by the HPA axis response.

The effects of acute stress on dopamine (28, 41) are the most obvious candidate for a mechanism by which stress might affect either model-based [via PFC (4)] or model-free [via striatum (19)] learning. Although the early sympathetic nervous system component of the stress response is known to result in rapid release of catecholamines in the PFC and other areas, and the resulting increase of dopamine (DA) levels is deleterious to PFC-dependent functions such as WM maintenance (7), our study focuses on the HPA axis stress response, for the simple, practical reason that the RL task takes time to administer. The release of glucocorticoids, indexed here by changes in cortisol levels, is observed to prolong this typically short-lived DA release in the PFC, among other regions (7, 41, 42). It is conceivable then that supraoptimal levels of DA (43) induced after the stressor and perpetuated by increases in cortisol release underlie the stress-induced deficits in central-executive–dependent, model-based behavior observed here.

In principle, a synergistic effect to weakening model-based learning might be strengthening model-free learning. Dopaminergic effects of stress might have been expected to produce such an effect as well, because increased striatal DA levels brought about by stress are hypothesized to increase overall sensitivity to reward (44). Although we found no such effect in our data, recent human probabilistic learning results may support this hypothesis (30). The probabilistic selection task used there does not formally dissociate model-free from model-based learning, but unlike our task, it does dissociate learning to choose vs. avoid—the locus of the reported stress effect.

Characterizing more precisely how neurophysiological stress response alters the expression of the two hypothesized valuations systems is of practical importance because acute stress is believed to facilitate drug-seeking and relapse (45). At the same time, prominent accounts of addiction (6) ascribe these compulsive behaviors to aberrant expression of the habitual system (instantiated here as the model-free system) at the expense of the goal-directed action (instantiated here as the model-based system). The finding that HPA axis response selectively reduces model-based contributions to behavior dovetails neatly with these accounts: perhaps the drug-seeking and/or relapse engendered by acute stress can be explained in part by a breakdown of the prospective, model-based valuation system.

Although the breakdown of top-down and prefrontal-dependent functions (7) is assumed to underlie the deleterious effects of neurophysiological stress response on model-based choice, a resource-allocation explanation of these results merits speculation here. One influential proposal suggests that people adapt to stress by falling back on strategies with fewer cognitive demands and in doing so, preventing unreliable performance that would ensue from failure to carry out more resource-demanding strategies (46). A recent, more computational proposal (20) frames the arbitration between model-based vs. model-free RL as a tradeoff between time cost and behavioral flexibility, both of which are high in model-based but low in model-free RL. Were the neurophysiological stress response to promote internal time pressure, we would expect the effects observed here. However, whether people register, implicitly or explicitly, the temporal and cognitive costs of model-based choice warrants future research.

Table 2. Logistic regression coefficients indicating the influence of cortisol response, outcome of previous trial, and transition type of previous trial, on response repetition

Coefficient	Estimate (SE)	*P* value
(Intercept)	1.84 (0.16)	<0.0001*
Reward	0.80 (0.09)	<0.0001*
Transition	0.10 (0.05)	0.060
Cortisol delta	0.06 (0.16)	0.712
Reward × transition	0.25 (0.06)	<0.0001*
Cortisol delta × reward	0.05 (0.09)	0.554
Cortisol delta × transition	–0.07 (0.06)	0.180
Cortisol delta × reward × transition	–0.14 (0.06)	0.018*

*Significance at the 0.05 level.

Materials and Methods

Participants. Fifty-six healthy individuals from the New York University community participated in this experiment (30 women, age: M = 25.67 y; SD = 7.27 y) and were paid 5 cents per rewarded trial to incentivize performance. The proportions of women in control and stress conditions were 0.50 and 0.58, respectively. All participants provided written informed consent in accordance with procedures approved by the New York University Committee on Activities Involving Human Subjects. Following previous work (21), we identified and excluded participants who failed to demonstrate engagement with the choice task. Specifically, we excluded data of four participants who failed to meet a response deadline greater than 15 times. We used a further step to remove participants who failed to demonstrate sensitivity to rewards in the decision task using second-stage choices, excluding the data of four participants who repeated previously rewarded second-stage responses—i.e., *P*(stay|win)—at a rate less than 50%.

Cortisol Measurement. To assess stress responses, saliva samples were collected throughout the task to assess cortisol concentrations. Samples were collected using an absorbent oral swab that participants placed under their tongues for 2 min. To control for diurnal rhythms in cortisol levels, all participants were run between the hours of 1:00 and 6:00 PM. Sample collection occurred at baseline after a 10-min acclimation period (s1), immediately after OSPAN measurement and task instructions (s2, ~25 min after s1), 10 min after CPT administration (s3, ~43 min after s1), and immediately following the RL task (s4, ~64 min after s1). Cortisol responses to stress were expected to peak during the RL task (10 min after the stress manipulation) (32). Samples were frozen and preserved immediately after testing at −30 °C and were transported frozen to a Clinical Laboratory Improvement Amendments-certified analytical laboratory where cortisol concentrations were determined with high-sensitivity enzyme immunoassay kits (Salimetrics). Duplicate assays were conducted for each sample interval, and the average of the two values was used in our analyses. Because of the skewed nature of cortisol concentration distributions, these values were log-transformed in all statistical tests (29). For each subject, cortisol delta was calculated by subtracting the average of s1 and s2 (pre-CPT) from the average of s3 and s4 (post-CPT).

OSPAN Measurement. To assess working memory capacity, we administered an automated version of the OSPAN procedure (47), which required participants to remember a series of letters while performing a series of arithmetic problems and which lasted ~15 min. OSPAN scores were calculated by summing the number of letters selected for all correctly selected sets and ranged from 11 to 75 (M = 48.08, SD = 17.61).

Stress Induction. In the stress condition, subjects were administered the CPT, described previously (33). Briefly, subjects in the stress condition were asked to immerse their right hand up to and including the wrist for 3 min in ice water (0–5 °C). Subjects in the control condition submerged their right hand up to and including the wrist for 3 min into room temperature water (21–30 °C). Immediately after, subjects indicated on a scale ranging from 0 (not at all) to 10 (very much) how unpleasant they found the immersion procedure.

RL Task. Immediately after the OSPAN procedure, participants were given the task instructions and completed 10 practice trials to familiarize themselves with the task structure and response procedure. Note that at this point, the control and stress groups were subject to the identical procedure, and thus differences in choice behavior could not be attributable to the conditions under which task instructions were given. Following administration of the cold pressor test and cortisol sample s3, participants completed 200 trials of the two-step RL task (Fig. 1*A*) immediately after sample s3 was taken. In the first step, two fractal images appeared on a black background (indicating the initial state), and there was a 1.5-s response window in which participants could choose the left- or right-hand response using the Z or ? key, respectively. After a choice was made, the selected action was highlighted for the remainder of the response window followed by the background color changing according to the second-stage state the participant had transitioned to. After the transition, the background color changed to reflect the second-stage state and the selected first-stage action moved to the top of the screen. Two fractal images, corresponding to the actions available in the second stage, were displayed, and participants again had 1.5 s to make a response. The selected action was highlighted for the remainder of the response window. Then, either a picture of a quarter was shown (indicating that they had been rewarded that trial) or the number zero (indicating that they had not been rewarded that trial) was shown. The reward probabilities associated with second-stage actions were governed by independently drifting Gaussian random walks (SD = 0.025) with reflecting boundaries at 0.25 and 0.75. The mapping of actions to stimuli and transition probabilities was randomized across participants.

Data Analysis. Cortisol deltas were log(+1) transformed to remove positive skew and were, along with OSPAN scores, entered into the RL model and regressions as z-scores. Details of the model fitting procedure and the regression specification are provided in *SI Text*. We fit subjects' choices using a full RL model that allows for choices to be influenced by the entire preceding history of rewards. The model follows closely the hybrid model described in ref. 16. For each parameter estimate, we computed a 95% CI; if 0 falls outside this interval, we can reject the null hypothesis that the true value is zero or more extreme with 95% confidence.

ACKNOWLEDGMENTS. This work was supported by a National Institute of Mental Health Grant (1R01MH087882-01 to N.D.D.), a National Institutes of Health Grant R01 AG039283 (to E.A.P.), and an Award in Understanding Human Cognition from the McDonnell Foundation (to N.D.D.).

1. Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* 8(12):1704–1711.
2. Kahneman D, Frederick S (2002) *Heuristics and Biases: The Psychology of Intuitive Judgment*, eds Gilovich T, Griffin D, Kahneman D (Cambridge Univ Press, Cambridge, UK), pp 49–81.
3. Sloman SA (1996) The empirical case for two systems of reasoning. *Psychol Bull* 119(1):3–22.
4. Balleine BW, O'Doherty JP (2010) Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35(1):48–69.
5. Redish AD, Jensen S, Johnson A (2008) A unified framework for addiction: Vulnerabilities in the decision process. *Behav Brain Sci* 31(4):415–437.
6. Everitt BJ, Robbins TW (2005) Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nat Neurosci* 8(11):1481–1489.
7. Arnsten AFT (2009) Stress signalling pathways that impair prefrontal cortex structure and function. *Nat Rev Neurosci* 10(6):410–422.
8. Leder J, Häusser JA, Mojzisch A (2013) Stress and strategic decision-making in the beauty contest game. *Psychoneuroendocrinology* 38(9):1503–1511.
9. Pabst S, Brand M, Wolf OT (2013) Stress and decision making: A few minutes make all the difference. *Behav Brain Res* 250:39–45.
10. Porcelli AJ, Delgado MR (2009) Acute stress modulates risk taking in financial decision making. *Psychol Sci* 20(3):278–283.
11. Putman P, Antypa N, Crysovergi P, van der Does WAJ (2010) Exogenous cortisol acutely influences motivated decision making in healthy young men. *Psychopharmacology (Berl)* 208(2):257–263.
12. Schwabe L, Wolf OT (2011) Stress-induced modulation of instrumental behavior: from goal-directed to habitual control of action. *Behav Brain Res* 219(2):321–328.
13. Starcke K, Wolf OT, Markowitsch HJ, Brand M (2008) Anticipatory stress influences decision making under explicit risk conditions. *Behav Neurosci* 122(6):1352–1360.
14. Frank MJ, Seeberger LC, O'reilly RC (2004) By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science* 306(5703):1940–1943.
15. Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16(5):1936–1947.
16. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69(6):1204–1215.
17. Gershman SJ, Markman AB, Otto AR (2012) Retrospective revaluation in sequential decision making: A tale of two systems. *J Exp Psychol Gen*, in press.
18. Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66(4):585–595.
19. Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275(5306):1593–1599.
20. Keramati M, Dezfouli A, Piray P (2011) Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLOS Comput Biol* 7(5):e1002055.
21. 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(5):751–761.
22. Lupien SJ, Gillin CJ, Hauger RL (1999) Working memory is more sensitive than declarative memory to the acute effects of corticosteroids: A dose-response study in humans. *Behav Neurosci* 113(3):420–430.
23. Schoofs D, Wolf OT, Smeets T (2009) Cold pressor stress impairs performance on working memory tasks requiring executive functions in healthy young men. *Behav Neurosci* 123(5):1066–1075.
24. Qin S, Hermans EJ, van Marle HJF, Luo J, Fernández G (2009) Acute psychological stress reduces working memory-related activity in the dorsolateral prefrontal cortex. *Biol Psychiatry* 66(1):25–32.
25. Schwabe L, Wolf OT (2009) Stress prompts habit behavior in humans. *J Neurosci* 29(22):7191–7198.
26. Schwabe L, et al. (2007) Stress modulates the use of spatial versus stimulus-response learning strategies in humans. *Learn Mem* 14(1):109–116.
27. Anstrom KK, Woodward DJ (2005) Restraint increases dopaminergic burst firing in awake rats. *Neuropsychopharmacology* 30(10):1832–1840.
28. Abercrombie ED, Keefe KA, DiFrischia DS, Zigmond MJ (1989) Differential effect of stress on in vivo dopamine release in striatum, nucleus accumbens, and medial frontal cortex. *J Neurochem* 52(5):1655–1658.
29. Petzold A, Plessow F, Goschke T, Kirschbaum C (2010) Stress reduces use of negative feedback in a feedback-based learning task. *Behav Neurosci* 124(2):248–255.
30. Lighthall NR, Gorlick MA, Schoeke A, Frank MJ, Mather M (2013) Stress modulates reinforcement learning in younger and older adults. *Psychol Aging* 28(1):35–46.
31. Conway ARA, Kane MJ, Engle RW (2003) Working memory capacity and its relation to general intelligence. *Trends Cogn Sci* 7(12):547–552.
32. McRae AL, et al. (2006) Stress reactivity: Biological and subjective responses to the cold pressor and Trier Social stressors. *Hum Psychopharmacol* 21(6):377–385.
33. Lee MD (2011) How cognitive modeling can benefit from hierarchical Bayesian models. *J Math Psychol* 55(1):1–7.
34. Smittenaar P, Fitzgerald THB, Romei V, Wright ND, Dolan RJ (2013) Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans [published online ahead of print on October 22, 2013]. *Neuron*, 10.1016/j.neuron.2013.08.009.
35. Wunderlich K, Smittenaar P, Dolan RJ (2012) Dopamine enhances model-based over model-free choice behavior. *Neuron* 75(3):418–424.
36. Starcke K, Brand M (2012) Decision making under stress: A selective review. *Neurosci Biobehav Rev* 36(4):1228–1248.
37. Shamosh NA, et al. (2008) Individual differences in delay discounting: Relation to intelligence, working memory, and anterior prefrontal cortex. *Psychol Sci* 19(9):904–911.
38. Bickel WK, Jarmolowicz DP, Mueller ET, Gatchalian KM, McClure SM (2012) Are executive function and impulsivity antipodes? A conceptual reconstruction with special reference to addiction. *Psychopharmacology (Berl)* 221(3):361–387.
39. Peters J, Büchel C (2011) The neural mechanisms of inter-temporal decision-making: understanding variability. *Trends Cogn Sci* 15(5):227–239.
40. Stern Y (2009) Cognitive reserve. *Neuropsychologia* 47(10):2015–2028.
41. Butts KA, Weinberg J, Young AH, Phillips AG (2011) Glucocorticoid receptors in the prefrontal cortex regulate stress-evoked dopamine efflux and aspects of executive function. *Proc Natl Acad Sci USA* 108(45):18459–18464.
42. Nagano-Saito A, et al. (2013) Stress-induced dopamine release in human medial prefrontal cortex-(18) F-Fallypride/PET study in healthy volunteers. *Synapse* 67(12):821–830.
43. Cools R, D'Esposito M (2011) Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biol Psychiatry* 69(12):e113–e125.
44. Mather M, Lighthall NR (2012) Both Risk and Reward are Processed Differently in Decisions Made Under Stress. *Curr Dir Psychol Sci* 21(2):36–41.
45. Sinha R (2001) How does stress increase risk of drug abuse and relapse? *Psychopharmacology (Berl)* 158(4):343–359.
46. Steinhauser M, Maier M, Hübner R (2007) Cognitive control under stress: How stress affects strategies of task-set reconfiguration. *Psychol Sci* 18(6):540–545.
47. Unsworth N, Heitz RP, Schrock JC, Engle RW (2005) An automated version of the operation span task. *Behav Res Methods* 37(3):498–505.

Supporting Information

Otto et al. 10.1073/pnas.1312011110

SI Text

Reinforcement Learning Model. The task consists of three states (first stage: s_A; second stage: s_B and s_C), each with two actions (a_A and a_B). The hybrid model consists of model-based and model-free subcomponents, both of which estimate a state-action value function $Q_{MF}(s,a)$ (model-free) and $Q_{MB}(s,a)$ (model-based) that maps each state-action pair to its expected future reward. On trial t, we denote the first-stage state (always s_A) by $s_{1,t}$, the second-stage state by $s_{2,t}$, the chosen first- and second-stage actions by $a_{1,t}$ and $a_{2,t}$, and the first- and second-stage rewards as $r_{1,t}$ (always zero) and $r_{2,t}$.

Model-free component. For the model-free algorithm, we used State-Action-Reward-State-Action, SARSA(λ), temporal difference (TD) learning (1), which updates the value for the visited state-action pair at each stage i and trial t according to

$$Q_{MF}(s_{i,t},a_{i,t}) = Q_{MF}(s_{i,t},a_{i,t}) + \alpha\delta_{i,t},$$

where

$$\delta_{i,t} = \left[r_{i,t} + Q_{MF}(s_{i+1,t},a_{i+1,t})\right]/\alpha - Q_{MF}(s_{i,t},a_{i,t}),$$

is the reward prediction error (RPE), and α is a learning rate parameter. For the first-stage choice, $r_{1,t} = 0$ and the RPE is driven by the second-stage value, $Q_{MF}(s_{2,t},a_{2,t})$; conversely, at the second stage, we define $Q_{MF}(s_{3,t},a_{3,t}) = 0$, because there is no further value in the trial apart from the immediate reward $r_{2,t}$. Here we have rescaled the leading term in the reward prediction error by $1/\alpha$, relative to its usual definition (2, 3). Because this simply rescales the units of the Q values (by $1/\alpha^2$ and $1/\alpha$ at the first and second stage, respectively), the same data likelihoods are maintained via a corresponding rescaling of the first- and second-level inverse temperatures β_{MF} and β_2 in the choice rule below. This slight reparameterization facilitates group-level modeling by reducing the correlation of the β s with α.

The model uses an eligibility trace to propagate second-stage reward information to the first-stage values. Specifically, at the end of each trial, the first-stage values are updated according to

$$Q_{MF}(s_{1,t},a_{1,t}) = Q_{MF}(s_{1,t},a_{1,t}) + \lambda\delta_{2,t},$$

where λ is an eligibility trace decay parameter (4), and the omission of α (which would normally appear in this equation) again results from rescaling the update to match the scaling implied by the prediction error above. We assume that eligibility traces are reset to 0 between episodes (i.e., that eligibility does not carry over from trial to trial).

Additionally, at the end of each trial, we decayed the Q values for all of the nonchosen actions by multiplying them by $1-\alpha$ (5, 6). This decay makes the present model correspond more closely to the one-trial-back regression model described in the main text, in the limit as $\alpha \to 1$.

Model-based component. In general, a model-based reinforcement learning (RL) algorithm works by learning a transition function (mapping state-action pairs to a probability distribution over the subsequent state), and immediate reward values for each state, then computing cumulative state-action values by iterative expectation over these. Specialized to the structure of the current task, this amounts to, first, simply deciding which first-stage action maps to which second-stage state (because subjects were instructed that this was the structure of the transition contingencies), and second, learning immediate reward values for each of the second-stage actions (the immediate rewards at the first stage being always zero).

Following ref. 7, we modeled transition learning by assuming subjects simply chose between the two possibilities: $P(s_B|s_A,a_A) = 0.7$, $P(s_C|s_A,a_B) = 0.7$, or vice versa, $P(s_B|s_A,a_A) = 0.3$, $P(s_C|s_A,a_B) = 0.3$, with $P(s_B|s_A,a_B) = 1 - P(s_B|s_A,a_A)$ and $P(s_C|s_A,a_A) = 1 - P(s_C|s_A,a_B)$, according to whether more transitions had thus far occurred to s_B following a_A plus s_C following a_B, or vice versa to s_C following a_A plus s_B following a_B.

At the second stage (the only one where immediate rewards were offered), the problem of learning immediate rewards is equivalent to that for TD above, because $Q_{TD}(s_{2,t},a_{2,t})$ is just an estimate of the immediate reward $r_{2,t}$; with no further stages to anticipate, and the SARSA learning rule reduces to a delta rule for predicting the immediate reward. Thus, the two approaches coincide at the second stage, and we define $Q_{MB} = Q_{TD}$ at those states.

Finally the top level model-based values are defined from the transition and reward estimates using the Bellman Equation (8):

$$\begin{aligned} Q_{MB}(s_A,a_j) = {} & P(s_B|s_A,a_j) \max_{a\in\{a_A,a_B\}} Q_{MF}(s_B,a) \\ & + P(s_C|s_A,a_j) \max_{a\in\{a_A,a_B\}} Q_{MF}(s_C,a), \end{aligned}$$

where we have assumed these are recomputed at each trial from the current estimates of the transition probabilities and rewards.

Choice rule. Finally, to connect the values to choices, we use a softmax choice rule, which assigns a probability to each action according to the combination of both Q_{MB} and $Q_{MF,}$ each weighted with a separate inverse temperature parameter, and β_{MB} and β_{MF}, which allow the two values to combine independently in determining first-stage choice. (Note that this is algebraically equivalent to the formulation used in ref. 7, under the substitution $\beta_{MB} = w\beta$ and $\beta_{MF} = (1-w)\beta$. This change of variables again facilitates group level modeling of individual differences in the influence of either system.

The probability of a choice at the first stage is calculated, accordingly, as

$$P(a_{i,t} = a|s_{1,t}) = \frac{\exp\left[\beta_{MB}\cdot Q_{MB}(s_{1,t},a) + \beta_{MF}\cdot Q_{MF}(s_{1,t},a) + p\cdot rep(a)\right]}{\sum_{a'}\exp\left[\beta_{MB}\cdot Q_{MB}(s_{1,t},a') + \beta_{MF}\cdot Q_{MF}(s_{1,t},a') + p\cdot rep(a')\right]}.$$

The indicator function $rep(a)$ is defined as 1 if a is a top-stage action and is the same one as was chosen on the previous trial, zero otherwise. Together with the “stickiness” parameter p, this captures first-order perseveration ($p > 0$) or switching ($p < 0$) in the first-stage choices. Second-stage choices are modeled with an analogous but simpler softmax rule, with only a single value term $Q_{MF}(s_{2,t}, a)$, with its own inverse temperature β_2 and omitting the $rep(a)$ term.

Group-level modeling. The foregoing describes the modeling of a single subject's data. This model was embedded within a multilevel random effects model to estimate it for all subjects simultaneously. All of the free parameters of the model (α, λ, β_{MB}, β_{MF}, β_2, p) were taken as random effects, instantiated separately for each subject s from a common group level distribution. For parameters with infinite support, the group level distributions were Gaussian with free mean and SD

$$\beta_{2_s} \sim N(\mu_{\beta 2}, \sigma_{\beta 2}),$$

and similarly for p_s. To test the dependence of the model-based and model-free effects on cortisol and Operation Span (OSPAN), these effects and their interaction were entered into a regression at the group level

$$\beta_{MB_s} \sim N\Big[\mu_{\beta MB} + \beta_{MBcort}\, cort(s) + \beta_{MBospan}\, ospan(s) + \beta_{MBcxo}\, cort(s) \cdot ospan(s)\Big],$$

and similarly for β_{MF_s}. Accordingly, nonzero values of the slopes β_{MBcort}, $\beta_{MBospan}$, and β_{MBcxo} signify correlations between cortisol delta, OSPAN, and the interaction between the two, analogous to the covariate effects tested in the logistic regression in the main text.

The parameters with support in $[0, 1]$ were assumed to be drawn from a group-level beta distribution

$$\alpha_s \sim Beta(A_\alpha, B_\alpha)$$

and similarly for λ_s.

Finally, we estimated the parameters of the group level distributions ($\mu_{\beta 2}$, etc.) using uninformative priors: for all means, the broad Gaussian $N(0, 100)$, for all SDs, the heavy-tailed $Cauchy(0, 2.5)$. Finally, our priors for the A and B parameters of the beta distributions were given using a change of variables that characterizes the distribution's mean $M = \frac{A}{A+B}$ and spread $S = \frac{1}{\sqrt{a+b}}$, the latter approximating its SD. This allowed us to take as uninformative hyperpriors the uniform distributions $M \sim U(0, 1)$ and $S \sim U(0, \infty)$ (the latter improper) (9).

Estimation. We estimated the joint distribution of the parameters of the model, conditional on all subjects' observed choices and rewards. For this, we used Markov Chain Monte Carlo (MCMC) techniques (specifically the No-U-Turn variant of Hamiltonian Monte Carlo) as implemented in the Stan modeling language (10). Given a probabilistic generative model (the above equations) and a subset of observed variables, MCMC techniques provide samples from the conditional joint distribution over the remaining random variables. We ran four chains of 2,000 samples each, discarding the first 1,000 samples of each chain for burn-in. We examined the chains visually for convergence and also computed Gelman and Rubin's (11) potential scale reduction factors. For this, large values indicate convergence problems, whereas values near 1 are consistent with convergence. We ensured that these diagnostics were less than 1.1 for all variables.

Results. Table S2 reports the free parameters of the model by their group-level means and variances over individual subjects. Also reported are the regression slopes estimating how individuals' parameter settings covaried with cortisol deltas, OSPAN scores, or their interaction. This uncertainty is reported via quartiles: the median and 25th and 75th percentiles of the distribution. Of note, the group-level mean α was centered on 0.34, characteristic of a more gradual (and thus, less recency driven) learning process than is ascribed by the regression analysis in the main text, which assumes a learning rate of 1 (that is, only the most recent trial influences choice), supporting the conclusion that our reported effects apply to longer-term incremental learning, and are not limited to short-term patterns of win-stay-lose-shift adjustments.

Regression Analysis. We specified a mixed-effects logistic regression to explain the first-stage choice on each trial t (coded as stay vs. switch) using binary predictors indicating if reward was received on t-1 and the transition type (common or rare) that had produced it. Logistic regressions were conducted as mixed-effects models, performed using the lme4 package (12) in the R programming language. Within-subject factors (the intercept, main effects of reward and transition, and their interaction) were taken as random effects across subjects, and estimates and statistics reported are at the population level. Individual model-based and model-free effect sizes (the model-based and model-free indices used in Figs. S1 and S2) were calculated from posterior estimates, conditional on the estimated top-level effects. Planned contrasts were conducted using the esticon function (package doBy) (13) on the estimated model.

As an initial examination, we estimated a model that included both experimental condition (stress v.s control) and cortisol delta as between subjects-factors (Table S3). Statistically, we found a significant negative interaction between cortisol response (quantified by cortisol delta; *Materials and Methods*), previous reward, and transition type ($P < 0.01$), confirming that cortisol response effectively attenuated the model-based signature of choice. Experimental condition (stress vs. control), however, did not exert significant influence on choice-related variables, nor did it significantly interact with the interaction between cortisol response and these trial-by-trial variables. That cortisol response yields greater explanatory leverage on behavior than experimental condition mirrors the results of recent examinations of stress and decision-making (14, 15). A separate regression, excluding condition, is reported in Table 2. Further, this regression confirmed that cortisol response did not influence the simple effect of previous reward—the hallmark of model-free learning ($P > 0.5$). Moreover, the effect of cortisol response on model-based contributions trended larger than model-free contributions (linear contrast between the reward effect and the reward × transition interaction, $P = 0.07$), positively demonstrating the selectivity of the effect to model-based RL and suggesting that cortisol response does not merely bring about a generalized decline in performance.

To visualize the relationship between cortisol response and model-based contribution to behavior analogously to the computational model weights, we computed for each subject a model-based index (the individual's coefficient estimate for the previous reward × transition type interaction as in Fig. 5*A*, the marker of model-based updating). Fig. S1*A* plots the model-based index as a function of cortisol response and condition, suggesting that the model-based contribution to choice decreased as a function of cortisol increase. Plotting the model-free index, an individual measure of the model-free contribution to choice (the coefficient for the main effect of the previous trial's reward on choice), as a function of cortisol response and condition revealed no apparent attenuation of model-free choice by cortisol response or condition (Fig. S1*B*).

We applied the same analysis approach to examine how individual working-memory (WM) capacity—operationalized by OSPAN—modulates the effect of cortisol response on model-based choice. Accordingly, we examined this relationship with a logistic model examining how cortisol delta and OSPAN interacted with the same trial-by-trial variables in the above analysis (previous reward and transition type; see Table S4 for full model specification and coefficient estimates). Critically, OSPAN significantly interacted with the three-way interaction between cortisol response, previous reward, and previous transition type (the interaction signifying cortisol response's effect on model-based choice, $P < 0.01$). This relationship is visualized in Fig. S2, analogous to Fig. 4: among subjects low in WM capacity, cortisol delta reduced the expression of model-based choice (Fig. 4*A*), but among subjects high in WM capacity, cortisol response did not produce an appreciable impact on model-based contributions to behavior (Fig. 4*B*).

1. Rummery GA, Niranjan M (1994) *On-Line Q-Learning Using Connectionist Systems* (Cambridge Univ, Cambridge, UK).
2. Camerer C, Ho T-H (1999) Experienced-weighted attraction learning in normal form games. *Econometrica* 67(4):827–874.
3. Den Ouden HEM, et al. (2013) Dissociable Effects of Dopamine and Serotonin on Reversal Learning. *Neuron* 80:1090–1100.
4. Sutton RS, Barto AG (1998) *Reinforcement Learning* (MIT Press, Cambridge, MA).
5. Lau B, Glimcher PW (2005) Dynamic response-by-response models of matching behavior in rhesus monkeys. *J Exp Anal Behav* 84(3):555–579.
6. Ito M, Doya K (2009) Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *J Neurosci* 29(31):9861–9874.
7. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69(6):1204–1215.
8. Bellman R (1957) *Dynamic Programming* (Princeton Univ Press, Princeton, NJ).
9. Gelman A, Carlin JB, Stern HS (1995) *Bayesian data analysis* (Chapman & Hall, London).
10. Stan Development Team (2013) *Stan: A C++ Library for Probability and Sampling, Version 1.3.* Available at http://mc-stan.org/. Accessed November 20, 2013.
11. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* 7(4):457–472.
12. Pinheiro JC, Bates DM (2000) *Mixed-Effects Models in S and S-PLUS* (Springer, New York).
13. Højsgaard S, Halekoh U (2009) *doBy: Groupwise Computations of Summary Statistics, General Linear Contrasts and Other Utilities.* Available at http://CRAN.R-project.org/package=doBy. Accessed November 20, 2013.
14. Starcke K, Polzer C, Wolf OT, Brand M (2011) Does stress alter everyday moral decision-making? *Psychoneuroendocrinology* 36(2):210–219.
15. Leder J, Häusser JA, Mojzisch A (2013) Stress and strategic decision-making in the beauty contest game. *Psychoneuroendocrinology* 38(9):1503–1511.

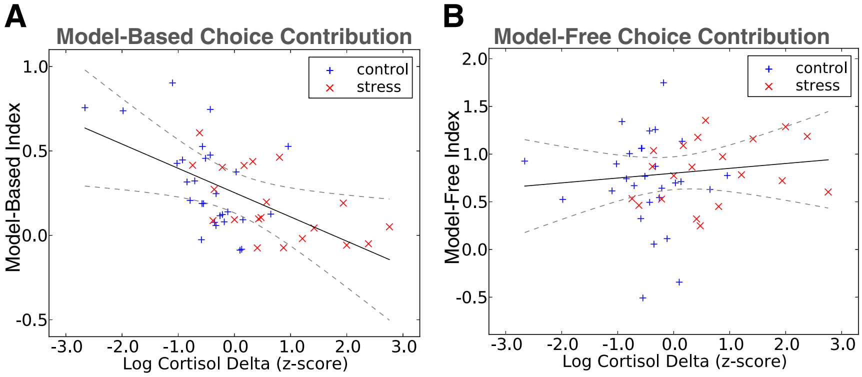

Fig. S1. Effect of cortisol response on model-based vs. model-free behavioral contributions. (*A*) Individual subjects' model-based effect sizes (arbitrary units) conditional on the group-level mixed-effects logistic regression, plotted separately for subjects in the control and stress conditions. The regression line is computed from the group-level log-linear effect of cortisol delta. There was a significant negative effect of cortisol delta on expression of model-based choice ($P < 0.05$), indicating cortisol change diminished its behavioral expression. (*B*) Subject-level effect-sizes for the model-free contribution to behavior. Note that there was no significant effect of cortisol change on expression of model-free choice ($P = 0.54$), indicating that expression of model-free contribution is spared. Dashed gray lines indicate 2 SEs, estimated from the group-level mixed effects regression.

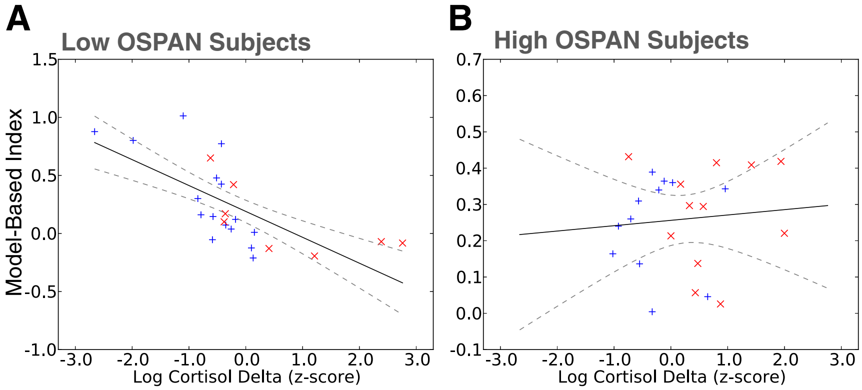

Fig. S2. Effect of cortisol response on expression of model-based behavior as a function of individual WM capacity as measured by OSPAN. Individual subjects' model-based effect sizes (arbitrary units) are plotted for low OSPAN subjects (*A*) and high OSPAN subjects (*B*). Cortisol response markedly dampened expression of model-based choice in the low OSPAN subgroup but not in the high OSPAN subgroup. Regression lines are computed from the group-level log-linear effect of cortisol delta for each subgroup. Dashed lines indicate 2 SEs, estimated from the group-level mixed effects regression.

Table S1. Mean cortisol response by group and sample time

	Sample			
Condition	t1 (baseline)	t2 (post-OSPAN)	t3 (post-CPT)	t4 (post-RL task)
Control (n = 28)	6.03 (3.08)	5.42 (2.47)	4.79 (2.31)	4.64 (2.65)
Stress (n = 20)	5.09 (3.54)	5.20 (3.01)	9.06 (6.49)	11.20 (13.26)

Salivary concentrations reported in nmol/L and are non–log transformed for interpretability.

Table S2. Group level estimates for the free parameters of the RL model and estimated slopes for the covariates

	Group-level means					
Percentile	β_{MB}	β_{MF}	p	β_2	λ	α
25	0.252	0.642	1.294	1.345	0.964	0.32
50	0.313	0.693	1.406	1.404	0.978	0.341
75	0.37	0.757	1.511	1.475	0.989	0.362
	Group-level variances					
	β_{MB}	β_{MF}	p	β_2	λ	α
25	0.389	0.458	0.922	0.590	0.016	0.193
50	0.436	0.504	0.996	0.643	0.038	0.206
75	0.496	0.554	1.077	0.711	0.069	0.221
	Covariate slopes					
	β_{MBcort}	$\beta_{MBospan}$	β_{MBcxo}	β_{MFcort}	$\beta_{MFospan}$	β_{MFcxo}
25	−0.285	0.051	0.318	−0.05	−0.061	−0.058
50	−0.226	0.113	0.42	0.007	−0.001	0.021
75	−0.163	0.17	0.525	0.063	0.062	0.108

For each parameter, the median posterior estimate is given, together with the quartiles of the posterior distribution. Note that the quartiles represent the width of uncertainty about the parameters' values (analogous to SEM), whereas the variances are estimates of the variability in the parameter estimates across the group of subjects.

Table S3. Logistic regression coefficients indicating the influence of cortisol response, stress condition, outcome of previous trial, and transition type of previous trial, on response repetition

Coefficient	Estimate (SE)	P value
(Intercept)	1.76 (0.20)	<0.0001*
Reward	0.72 (0.10)	<0.0001*
Transition	0.08 (0.07)	0.291
Cortisol delta	0.17 (0.33)	0.927
Condition	−0.11 (0.18)	0.690
Reward × transition	0.28 (0.07)	0.002*
Cortisol delta × reward	0.10 (0.18)	0.946
Cortisol delta × transition	0.02 (0.13)	0.170
Condition × reward	−0.09 (0.10)	0.702
Transition × cortisol delta	0.03 (0.07)	0.867
Condition × cortisol delta	0.20 (0.33)	0.391
Cortisol delta × reward × transition	−0.37 (0.13)	0.006*
Condition × reward × transition	0.11 (0.07)	0.321
Condition × cortisol delta × reward	0.18 (0.18)	0.314
Condition × cortisol delta × transition	0.06 (0.13)	0.347
Condition × cortisol delta × reward × transition	0.11 (0.13)	0.245

Critically, the cortisol delta × reward × transition was significant in the negative direction, indicating that cortisol response tempered model-based contribution to choice.

*Significance at the 0.05 level.

Table S4. Logistic regression coefficients indicating the influence of Operation Span (OSPAN) cortisol response, outcome of previous trial, and transition type of previous trial, on response repetition

Coefficient	Estimate (SE)	*P* value
(Intercept)	1.87 (0.17)	<0.0001*
Reward	0.77 (0.09)	<0.0001*
Transition	0.01 (0.05)	0.885
Cortisol delta	0.00 (0.16)	0.994
OSPAN	0.21 (0.17)	0.226
Reward × transition	0.20 (0.06)	<0.0001*
Cortisol delta × reward	0.03 (0.09)	0.734
Cortisol delta × transition	−0.08 (0.05)	0.090
OSPAN × reward	0.08 (0.09)	0.392
Transition × cortisol delta	0.08 (0.04)	0.084
OSPAN × cortisol delta	−0.12 (0.24)	0.633
Cortisol delta × reward × transition	−0.17 (0.06)	0.004*
OSPAN × reward × transition	0.09 (0.06)	0.099
OSPAN × cortisol delta × reward	0.06 (0.13)	0.619
OSPAN × cortisol delta × transition	0.35 (0.07)	<0.0001*
OSPAN × cortisol delta × reward × transition	0.23 (0.09)	0.009*

*Significance at the 0.05 level.