Acute stress impairs performance in a computationally hard cognitive task

Karlo Doroc[1] (karlo.doroc@unimelb.edu.au)

Nitin Yadav[1] (nitin.yadav@unimelb.edu.au)

Carsten Murawski[1] (carstenm@unimelb.edu.au)
[1]Centre for Brain, Mind and Markets, Faculty of Business and Economics, University of Melbourne
198 Berkeley Street, Carlton, VIC, 3041, Australia

Abstract

Acute stress triggers a cascade of physiological and psychological changes including heightened cortisol levels, perspiration, and anxiety. Existing research has focused on acute stress's effect on cognition in basic tasks of executive functioning, but its effect on computationally harder tasks is not well understood. Here, in a within-participants laboratory experiment (n=42, mostly college students), we test for an effect of acute stress on decision-making at varying levels of computational hardness in the 0-1 Knapsack Decision Problem. We find that acute stress, induced via the Trier Social Stress Test, leads to impaired decision quality irrespective of the level of computational hardness. Among cortisol responders, higher cortisol levels were associated with lower decision quality and higher time on task. Our findings help bridge the gap between research on executive functioning tasks and 'real-world decisions', building a more nuanced understanding of how acute stress affects decision-making.

Keywords: Acute stress; decision-making; computational hardness; complexity; experimental psychology

Introduction

Stress is a ubiquitous experience occurring when a situation's demands exceed an organism's regulatory capacity (Starcke & Brand, 2012). Acute, or 'in the moment' stress, triggers substantial physiological and psychological changes (Sapolsky, Romero, & Munck, 2000). Physiologically, the sympathetic-adrenal-medullary (SAM) axis activates the fight-or-flight response, with the likes of heart rate and perspiration increasing within seconds of stress onset and returning to baseline within minutes post-stressor (Starcke & Brand, 2012). In parallel, the hypothalamic-pituitary-adrenal (HPA) axis instigates a slower stress response, releasing glucocorticoids, such as cortisol, to mobilise energy stores and inhibit non-essential functions (Rodrigues, LeDoux, & Sapolsky, 2009). Cortisol levels increase within minutes of stress onset and remain elevated for 40-60 minutes post-stressor (Starcke & Brand, 2012). Psychologically, acute stress can be accompanied by states such as heightened alertness or anxiety (Grace et al., 2022; Sandi, 2013).

Behavioural studies show decision-making under acute stress is often impaired, but there is substantial variation. Most studies tend to focus on basic cognitive tasks, such as those assessing the three domains of executive functioning (EF): working memory, inhibition, and cognitive flexibility. There is a growing consensus that acute stress impairs working memory (Schoofs, Preuß, & Wolf, 2008; Shields, Bonner, & Moons, 2015) cognitive flexibility, and planning (Johnson, Dariotis, & Wang, 2012), while enhancing inhibition (Shields et al., 2015) and learning from negative feedback or stress-relevant contexts (Schwabe, Wolf, & Oitzl, 2010; Yu, 2016).

Similar trends have been found in other popular task paradigms, including risky (Duque, Cano-López, & Puig-Pérez, 2022) and two-stage choice tasks (Otto, Raio, Chiang, Phelps, & Daw, 2013), while evidence is mixed in domains such as choice consistency (Nitsch, Sellitto, & Kalenscher, 2021). Given the mixed results and the importance of often task-specific moderating variables, it is difficult to generalise from these studies to higher-order decision-making. We hypothesise that one major confounding factor of existing studies that may explain the mixed results is the lack of control for the computational hardness of existing tasks.

In our study, we address the issue. Our central contribution is to study the effect of acute stress on the quality of decisions at different levels of computational hardness. We have designed a task that allows us to quantify and manipulate computational hardness of decisions, that is, the amount of time and memory needed to decide correctly (Cheeseman, Kanefsky, & Taylor, 1991), in a precise and theoretically sound way. Our task is based on the 0-1 knapsack decision problem (KP). It recruits several cognitive processes, such as working memory and cognitive flexibility, and it can represent any decision scenario involving costs and benefits: from grocery shopping to investment management (Murawski & Bossaerts, 2016). However, our theoretical framework is independent of the specific task used. Prior work has shown the our measures of complexity used here affect human behaviour in the KP (Franco, Yadav, Bossaerts, & Murawski, 2021; Yadav, Murawski, Sardina, & Bossaerts, 2018), as well as tasks requiring abilities as diverse as spatial navigation and propositional logic (Franco, Doroc, Yadav, Bossaerts, & Murawski, 2022). Thus, our use of computational problems does not simply apply stress research to a novel setting. Rather, it is an altogether *more appropriate* setting, relative to alternatives such as EF or probabilistic tasks, for cognitive scientists interested in measuring and manipulating a decision's computational hardness in a generalisable, theoretically sound, and empirically validated way. Our design allows us to modulate both the computational resource requirements of the decision environment, via these complexity measures, and decision-makers' stress levels, via the Trier Social Stress Test (TSST).

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

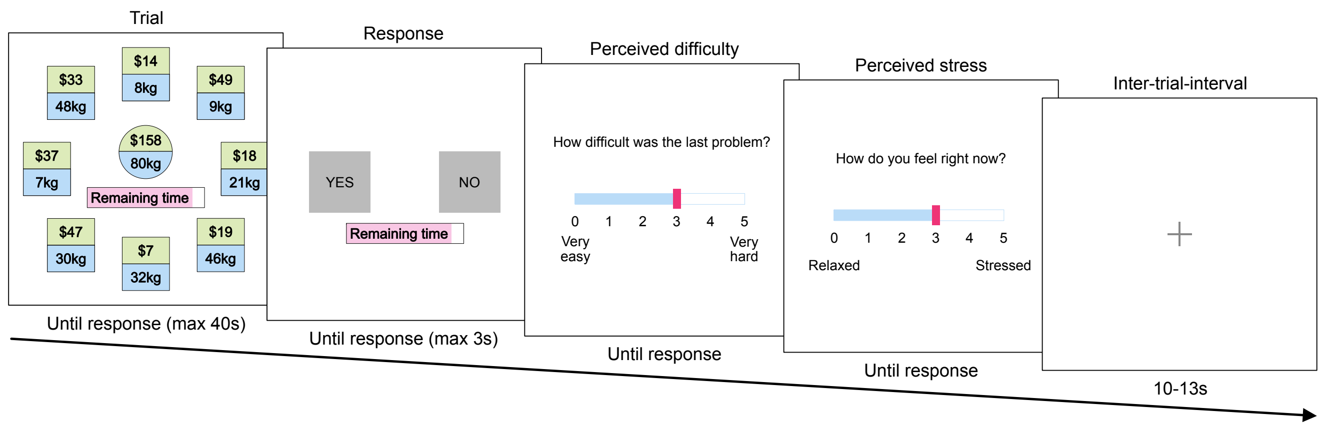

Figure 1: **Experimental task. Trial**: A set of 8 items is shown, each with a value and weight. A capacity constraint, target profit, and timer are shown at the screen's center. The goal is to decide whether there exists a subset of items for which (1) the sum of weights is lower or equal to the capacity constraint; and (2) the sum of values is at least the target profit. Nothing on the screen could be moved, selected or highlighted. **Response**: Participants selected YES or NO as their solution.

We expected that acute stress would have a more detrimental effect on decision quality at higher levels of computational hardness, that is, in situations in which the computational resource constraints are more binding.

Methods

Participants

The experimental protocol was approved by the University of Melbourne Human Research Ethics Committee (Ethics ID 23412). We recruited 42 participants (21 female, mean age 21.5 with standard deviation (SD) 3.6), predominantly college students, for an in-person laboratory experiment. Of the female participants, 16 were in the luteal phase, 3 in the follicular phase, and 2 in the menstrual phase of their menstrual cycle. To avoid anticipatory stress the study was described as assessing "physiological and emotional responses to performance and decision-making tasks" (Wemm & Wulfert, 2017). Each participant satisfied the standard exclusion criteria (Shields, Trainor, Lam, & Yonelinas, 2016) and did not eat, smoke, exercise, or drink anything besides water within two hours of the session. Participants were paid a A$10 show-up fee and A$0.7 for each correct response in the knapsack task. The average total payment was A$98.8 per participant.

Materials

Physiological and psychological measurements We collected three primary stress measures: five salivary cortisol samples (HPA-axis response), five samples of the Positive Affect Negative Affect Schedule (PANAS; subjective response) (Watson, Anna, & Tellegen, 1988), and electrodermal activity (EDA), collected with a Shimmer3+ device fitted on participants' non-dominant hand, sampled at 64 Hz (SAM response). Further, every 9 trials the participants rated how they were feeling on a scale from 0 (relaxed) to 5 (stressed), which we denote perceived stress (subjective response). Eye-tracking data were collected, but are not reported here.

Experimental task: the 0-1 Knapsack Decision Problem An instance of the KP consists of a set of items $I = 1, \ldots, N$ with weights $w_1, \ldots, w_N$ and values $v_1, \ldots, v_N$, and two positive numbers c and p denoting the capacity and target profit of the knapsack. The problem is to decide, yes or no, whether there exists a set $S \subseteq I$ such that $\sum_{i \in S} w_i \leq c$ and $\sum_{i \in S} v_i \geq p$; see Figure 1 for more details. All weights, values, capacities, and target profits used were integers.

In each session participants completed 72 unique KP trials (see Figure 1). The order of trials, and the locations of items and response buttons on a given trial, were randomised for each participant. Each trial had two measures of performance: accuracy ($\{0, 1\}$) and time on task, operationalised as the proportion of the maximum available time used ([0,1]).

Instance sampling Instances were sampled based on their instance complexity (IC), as defined in Franco et al. (2021). IC measures the amount of computational resources an algorithm requires to solve an instance, where harder instances require more computational resources to solve. Formally, IC is the normalised distance between the target profit, p, and the maximum attainable profit, p^*: $IC = |\frac{p-p^*}{\sum v_i}|$

IC was used to measure computational hardness for three reasons. First, IC is highly correlated with Typical Case Complexity (TCC), a theoretically grounded, task-independent measure of computational hardness (Cheeseman et al., 1991) which has been shown to empirically affect behaviour in the knapsack task and other computationally hard problems (Franco et al., 2022). Second, IC is a more precise measure for the computational hardness of individual instances. While TCC is defined *on average* in relation to a random ensemble of instances, IC is defined precisely for an individual instance (Franco et al., 2021). Third, IC predicts performance equally well on satisfiable (solution is yes) and unsatisfiable (solution is no) instances, unlike more intuitive measures like the number of item combinations that satisfy the constraints or whether choosing the highest value-to-weight ratio items leads to the correct solution.

The value of IC is inversely proportional to the difficulty of the instance, and we randomly sampled 36 hard instances, $IC \in [|0.0268|, |0.0321|]$, and 36 easy instances, $IC \in [|0.0971|, |0.105|]$. These parameters were chosen following pilot testing to target an average accuracy of 90% for easy instances and 70% for hard instances. For both easy and hard instances, half of the instances were satisfiable (unsatisfiable). The task was programmed in Unity3d (version 2020.3.28f1) and administered on a desktop computer.

Procedure

Participants completed a stress and a control session on separate days, on average 4 days apart, but at the same time of day. Sessions began either at 1pm or 4pm as natural cortisol levels are relatively stable in the afternoon (Allen, Kennedy, Cryan, Dinan, & Clarke, 2014). The location, experimenter, equipment, procedure, and tasks were identical across sessions, except for the stress manipulation. In the stress session, participants underwent the TSST (Kirschbaum, Pirke, & Hellhammer, 1993), considered the gold standard for eliciting acute stress responses in the laboratory (Bali & Jaggi, 2015). In the control session, participants underwent the placebo TSST (Het, Rohleder, Schoofs, Kirschbaum, & Wolf, 2009). Session order was counterbalanced across participants and sex.

The timeline of a typical session is detailed in Figure 2 and in total a session lasted up to 150 minutes. To improve salivary cortisol data reliability, participants were administered with 30g of dextrose powder dissolved in 200ml of water immediately prior to the rest period (Labuschagne, Grace, Rendell, Terrett, & Heinrichs, 2019). Once consumed, participants were administered with 100ml of plain water. After the second session, participants were debriefed by the experimenter, including a pre-recorded video where the confederates in their stress session explained the purpose of the TSST.

Stress manipulation

The TSST procedure was primarily adapted from the canonical version (Kudielka, Wüst, Kirschbaum, & Hellhammer, 2007), as well as from a recent guide (Labuschagne et al., 2019). The anticipatory, speech, and arithmetic phases each lasted 5 minutes. The speech topic was "imagine you are applying for your dream job, persuade the hiring panel to hire you". In the placebo TSST, it was "What are your hobbies?". The TSST arithmetic task was counting down from 2047 in steps of 17. In placebo, it was counting from 0 in steps of 15.

Data analysis pipeline

Salivary cortisol Salivary cortisol was analysed by Royal Melbourne Hospital Pathology using a Roche Cobas instrument and the Cobas e411 analyser (Chiu, Collier, Clark, & Wynn-Edwards, 2003; Vogeser et al., 2017). Participants were excluded from analysis if their baseline salivary cortisol levels were more than 3 SDs from the mean (Sandner, Zeier, Lois, & Wessa, 2021). Participants were classed as cortisol responders if their cortisol level exceeded baseline by at least 1.5 nmol/L (Miller, Plessow, Kirschbaum, & Stalder, 2013; Vogeser et al., 2017). As the distribution of cortisol values was highly skewed, a natural log transformation was applied. Our main stress measure, which has been used in several other stress and decision-making studies (Lenow, Constantino, Daw, & Phelps, 2017; Lighthall, Gorlick, Schoeke, Frank, & Mather, 2013; Otto et al., 2013), was:

$$\log \text{cort}\,\Delta = \frac{\log \text{cort}_3 + \log \text{cort}_4 + \log \text{cort}_5}{3} - \frac{\log \text{cort}_1 + \log \text{cort}_2}{2} \quad (1)$$

An exploratory cortisol measure, log predicted cortisol, was constructed to approximate the cortisol level on a given trial. To compute this we used cortisol samples 2-5 and regressed cortisol onto time using a loess regression. The predicted cortisol level at the time a trial began was stored. As loess regressions cannot extrapolate, for trials after the final cortisol sample we estimated a linear trend based on the last 3 loess predictions. We then applied a natural log transformation and z-scored the variable for analysis.

EDA An 8-minute window from both the rest period and the stress intervention was used to compute skin conductance level (SCL). The windows were identical for each participant, namely from T-10 to T-2 minutes, where T is the end of the period. Analysis was based on Taylor et al. (2015) and the corresponding code available on GitHub. The raw signal was passed through a 1Hz, fifth order, low-pass Butterworth filter. The median SCL for the rest period was treated as a baseline and subtracted from each point in the intervention period. The median of the baseline-adjusted intervention was then z-scored and used as a session-level variable. Upon visual inspection of the data, 12 participants' EDA data needed to be discarded due to recording quality problems.

Other measures Negative affect (NA) scores were square root transformed to reduce skewness. Mean change from baseline was then computed and z-scored for analysis.

Main analysis To account for individual differences in cortisol levels and task performance, we used hierarchical modelling with random effects on the intercept for each participant. We used three classes of regression models. Linear, logistic, and beta regressions were used to predict unbounded continuous variables, binary variables, and bounded continuous variables, respectively. If multiple models were considered the AIC was used for model selection. The main focus of the analysis was the full sample and cortisol responders. The independent variable used to assess stress was log cort Δ.

Behavioural analyses included complexity and satisfiability as binary variables, with 0 denoting easy (unsatisfiable) and 1 hard (satisfiable) instances, respectively. They capture separate features of the decision environment: computational hardness (complexity) and whether one must search for a solution or verify the absence of one (satisfiability).

After removing one cortisol outlier, we had 41 participants, of which 27 were classed as cortisol responders. To estimate

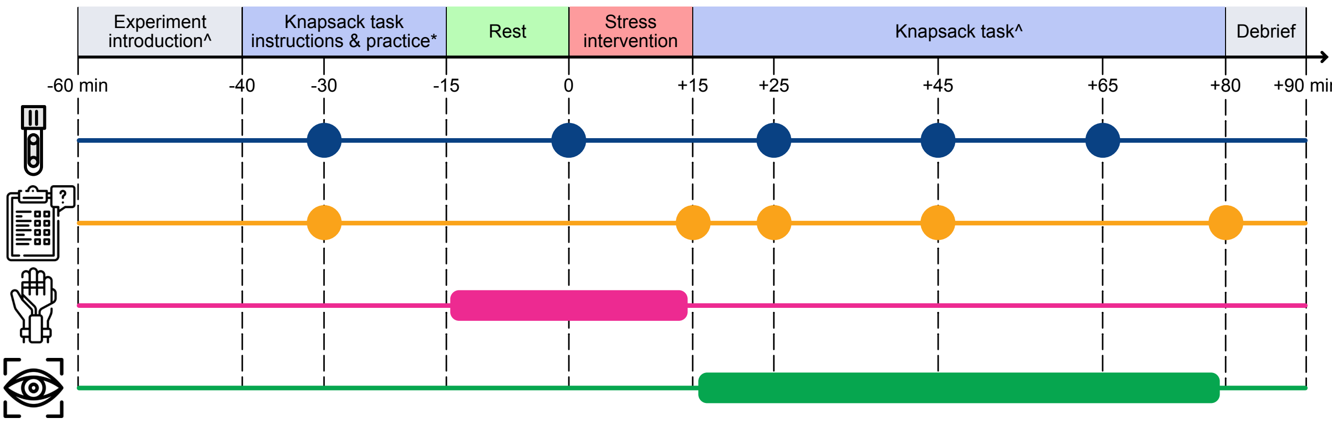

Figure 2: **Timeline of a typical experimental session.** Reading from top to bottom, the blue timeline corresponds to the salivary cortisol samples, orange to the PANAS, pink to SCL, and green to eye-tracking. The stress intervention was either the TSST (stress session) or placebo TSST (control session). All timings, except where noted, were identical across all participants and sessions. ˆSelf-paced with a fixed start time. *Self-paced with a fixed end time.

an upper bound on effect sizes, we computed marginal effects among the 6 participants, the 'high- stress' group, with significant increases in each of log cort Δ, NA, and SCL. EDA preprocessing was done in Python (version 3.9.12) and statistical analyses were done in R (version 4.1.2) using RStudio.

Results

Efficacy of acute stress manipulation

To verify that the acute stress manipulation was successful we performed paired t-tests for each stress measure collected. As shown in Table 1, all stress measures, but perceived stress, were higher in the stress condition than the control condition. This was the case for the full sample and cortisol responders.

To verify that log cort Δ was representative of a broader stress response we correlated it with our other stress measures. For the full sample we found significant correlations with the binary stress condition variable ($r = 0.45, p < 0.001$), log predicted cortisol ($r = 0.89, p < 0.001$), and the area under the cortisol response curve with respect to ground ($r = 0.52, p < 0.001$) and increase ($r = 0.95, p < 0.001$), respectively (Pruessner, Kirschbaum, Meinlschmid, & Hellhammer, 2003). There was no significant correlation with NA, perceived stress, or SCL. These results were qualitatively similar among cortisol responders.

To verify that the changes in cortisol persisted throughout the session, we estimated a linear mixed effects model regressing cortisol levels onto cortisol sample number, session number, session start time, sex, stress condition, and a stress condition $\times$ sample number interaction. Reporting the effects relevant to the stress condition, among the full sample we found a significant negative effect for sample number ($\beta = -0.26, CI_{95} = [-0.47, -0.04], p = 0.019$), a significant positive effect for a stress condition $\times$ sample number interaction ($\beta = 0.56, CI_{95} = [0.25, 0.86], p < 0.001$), and no main effect for stress condition. Among cortisol responders we found a significant and positive stress condition $\times$ sample number interaction ($\beta = 0.80, CI_{95} = [0.41, 1.18], p < 0.001$), with the respective main effects not significant. Thus, cortisol levels increased over time in the stress session, over and above any corresponding changes in the control session.

Table 1: Effect of acute stress manipulation.

Mean difference in stress measures (prior to z-scoring) between conditions for the full sample and cortisol responders. P-values computed via two-sided paired t-tests, except for SCL which was unpaired. AUCi and AUCg are the area under the cortisol response curve with respect to increase and ground, respectively. PA is positive affect. Per. stress is perceived stress. When comparing across stress conditions, $\dagger p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, NS $p > 0.1$.

	Full Sample		**Cort. Responders**	
Measure	Stress	Control	Stress	Control
log cort Δ	0.31^{***}	-0.11	0.56^{***}	-0.04
Cortisol	6.24^{***}	4.70	6.84^{***}	4.54
AUCi	117.26^{***}	-30.40	204.94^{***}	-6.48
AUCg	436.13^{***}	311.60	485.95^{***}	301.93
SCL	1.03^{***}	0.85	1.04^{***}	0.85
NA	14.58^{***}	12.98	14.9^{***}	12.9
PA	21.61^{*}	23.07	19.9^{**}	21.9
Per. stress	2.60^{NS}	2.55	$2.68^{\dagger}$	2.47

Effect of acute stress on behaviour

41 participants completed all 144 trials, with one completing 143. Averaged across sessions, mean accuracy was 78.3% (SD = 8.2%) for all trials, 88.1% (SD = 9.2%) for easy trials, and 68.5% for hard trials (SD = 10.5%). Mean time on task was 31.3 seconds (SD = 5.3) for all trials, 29.7 seconds (SD = 5.5) for easy trials, and 33 seconds (SD = 5.4) for hard trials.

Effect of acute stress on accuracy in the KP To test the effect of acute stress on accuracy we estimated a logistic mixed effects model regressing accuracy onto log cort Δ, complexity, a log cort Δ $\times$ complexity interaction, satisfiability, sex, trial number, and session number (AIC: 5,743.3). Time on task was omitted due to confounds with the complexity and stress measures. There was no significant effect for the interaction term or trial number and the AIC improved when these variables were removed (AIC: 5,740.6). For the full sample, we found a significant negative effect of complexity ($\beta = -1.26, CI_{95} = [-1.40, -1.12], p < 0.001$). The effect of log cort Δ was not significant ($\beta = -0.08, CI_{95} = [-0.16, 0.01], p = 0.069$). We also found significant positive effects for sex (males were coded as 1s) ($\beta = 0.40, CI_{95} = [0.12, 0.68], p = 0.004$) and session number (the second session was coded as 1s) ($\beta = 0.21, CI_{95} = [0.08, 0.34], p = 0.002$). There was no significant effect for satisfiability.

Among cortisol responders, there was no effect of sex (AIC: 3,644.7), and the AIC improved when this was omitted from the model (AIC: 3,644.4). We observed a significant positive effect for session number ($\beta = 0.20, CI_{95} = [0.03, 0.36], p = 0.019$) and negative effects for complexity ($\beta = -1.37, CI_{95} = [-1.55, -1.20], p < 0.001$), log cort Δ ($\beta = -0.13, CI_{95} = [-0.23, -0.03], p = 0.008$), and satisfiability ($\beta = -0.31, CI_{95} = [-0.47, -0.15], p < 0.001$). This implies a marginal effect whereby a one standard deviation increase in log cort Δ is associated with a 1.7 percentage point reduction in accuracy (see Figure 3). To estimate an upper bound on this effect size, we estimated a corresponding regression for the high-stress group. Log cort Δ was not significant for this group. Replacing log cort Δ with the binary stress variable yielded a significant negative effect ($\beta = -0.47, CI_{95} = [-0.81, -0.13], p = 0.008$). This translates into a marginal effect of a 7.2 percentage point reduction in accuracy under stress.

Effect of acute stress on time on task in the KP To examine the effect of acute stress on time on task we estimated a mixed-effects beta regression, regressing time on task onto log cort Δ, complexity, a log cort Δ $\times$ complexity interaction, satisfiability, sex, trial number, and session number (AIC: -29,499.5). There was no significant effect of the interaction term or sex and the AIC improved when these variables were removed (AIC: -29,503.4). For the full sample, we found a significant positive effect of complexity ($\beta = 0.43, CI_{95} = [0.37, 0.49], p < 0.001$) and log cort Δ ($\beta = 0.08, CI_{95} = [0.04, 0.12], p < 0.001$). We found significant negative effects of satisfiability ($\beta = -0.28, CI_{95} = [-0.33, -0.22], p < 0.001$), trial number ($\beta = -0.008, CI_{95} = [-0.009, -0.006], p < 0.001$), and session number ($\beta = -0.14, CI_{95} = [-0.19, -0.08], p < 0.001$).

The same model fit best among cortisol responders, where we observed significant and positive effects of complexity ($\beta = 0.47, CI_{95} = [0.40, 0.54], p < 0.001$) and log cort Δ ($\beta = 0.07, CI_{95} = [0.03, 0.11], p = 0.001$), while significant and negative effects of satisfiability ($\beta = -0.29, CI_{95} = [-0.36, -0.22], p < 0.001$), trial number ($\beta = -0.008, CI_{95} = [-0.010, -0.006], p < 0.001$), and session number ($\beta = -0.19, CI_{95} = [-0.26, -0.12], p < 0.001$). The effect size implies a marginal effect of a 0.8 percentage point increase in time on task (equivalent to 0.3 seconds) for a one standard deviation increase in log cort Δ. To estimate an upper bound on this effect size we estimated a corresponding regression for the high-stress group. Log cort Δ yielded a significant positive effect ($\beta = 0.24, CI_{95} = [0.11, 0.36], p < 0.001$). This translates into a marginal effect of a 2.6 percentage point (or 1.1 second) increase in time on task for a one standard deviation increase in log cort Δ.

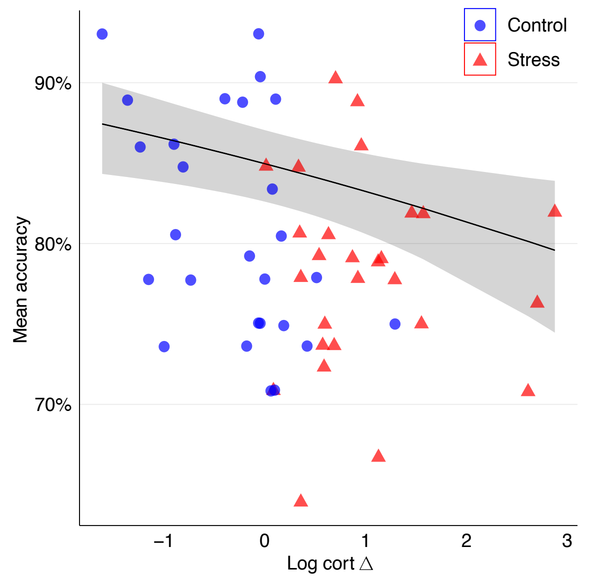

Figure 3: **Marginal effect of acute stress on accuracy**. Change in mean accuracy conditional on the mixed-effects logistic regression for cortisol responders, plotted as a function of log cort Δ. The regression line is computed from the group-level intercept and log cort Δ fixed effects. Shaded area shows the 95% confidence interval.

Exploratory results Post-hoc we repeated our earlier analysis and replaced log cort Δ with our trial-level measure, log predicted cortisol. The estimated models were otherwise identical to the main analysis. Among cortisol responders, there was a significant negative main effect ($\beta = -0.26, CI_{95} = [-0.44, -0.07], p < 0.006$) and no interaction effect with complexity. There was no significant main or interaction effect of log predicted cortisol on accuracy for the full sample, or on time on task for either sample.

Discussion

Our central finding is that, in response to an acute psychosocial stressor, higher levels of cortisol impaired performance on the KP, independent of the level of computational hardness. While others have argued that acute stress has an increasingly impairing effect on performance as the complexity of the task increases (Arnsten, 2009; Plieger & Reuter, 2020; Sandi, 2013), these arguments have been limited in three ways. Firstly, these studies tend to have a narrow focus on cognitive tasks that isolate specific executive functions. Secondly, these studies lack a formal framework with which to measure within-task task complexity in a task-independent manner. Thirdly, there is mixed evidence for the effects of stress among the tasks considered (Lai, Yeh, Lin, Hsu, & Wu, 2017; Long & Mo, 1971). Here, we offer the first evidence that acute stress impairs the quality of human decision-making at both low and high levels of computational hardness. We believe that our study helps to bridge the gap between basic EF tasks and real-world decisions.

Decision-making research across different disciplines has increasingly used a framework of 'cognitive resources' to describe the cognitive demands on decision-makers (Lieder & Griffiths, 2020; Shenhav et al., 2017). In this framework, constructs like energy, memory, and attention are cast as resources that are deployed in order to solve a problem or make a decision, with good decisions coming when the available cognitive resources meet or exceed those required by the task at hand. Along similar lines, the lens of a decision-maker engaging in information processing and/or computation has become widely adopted in cognitive science (Piccinini & Scarantino, 2010; Richards & Lillicrap, 2022). To the extent that stress can be considered a tax on cognitive resources (Vecchio, 1990), research on stress and computational problems such as the KP provide a promising avenue to test for a relationship between *cognitive resources* on the one hand and the *computational resource requirements* of the task environment on the other.

Our ability to use precise, theory-driven, and generalisable metrics of complexity at the trial level left us uniquely placed to test for an interaction effect between acute stress and complexity on performance. Surprisingly, we found no evidence for such an interaction. While we cannot rule out that a more precise measure of complexity would have captured an interaction effect, it would be surprising if IC could near perfectly capture the expected effect of complexity, but not the dimension of complexity relevant to acute stress. As we observed modest cortisol-related effect sizes, one possible explanation is a lack of statistical power. We also find no evidence for an inverse-U shaped relationship between cortisol and performance. Our data does not support the proposition that the acute stress response's increase of arousal levels trades off benefits from enhanced attention with costs to reasoning abilities, with an optimal level of arousal maximising performance (Plieger & Reuter, 2020; Yerkes & Dodson, 1908). Indeed, time on task results suggest that differences in effort and cognitive control may not explain the observed differences in behaviour, leaving open the mechanism by which acute stress affected decision-making performance in our experiment.

To our knowledge, we are the first to show that acute stress affects performance on a nondeterministic polynomial time complete (NP-complete) decision-making task. For our purposes, NP-complete problems have four key features: they are ubiquitous, they are computationally hard at the problem level but computational hardness varies considerably at the individual instance level, and NP-complete problems are mathematically related to each other, which suggests that findings in relation to one such problem are likely to generalise to all other NP-complete problems (Arora & Barak, 2009). Indeed, empirically, measures such as instance complexity, whilst first tested on humans in the KP, have been found to generalise to a diverse set of NP-complete problems requiring skills as varied as spatial navigation or propositional logic to solve (Franco et al., 2022). Examples of real-world NP-complete settings include scheduling, budgeting, route planning, and event planning. Thus, the implications of understanding how stress affects such complex cognitive processes are vast, with potential applications including improved design of work environments, educational programs, and therapeutic interventions.

Our study has several limitations. Data quality issues with the EDA data limited our effective sample size for EDA analyses. Our sample primarily consisted of undergraduate college students and, while our non-stress related results were consistent with other studies using the KP (Murawski & Bossaerts, 2016), we cannot be sure that the effects of acute stress observed here will generalise to other populations.

Future work could test the generality of our findings with extensions employing a between-subjects design, more refined computational complexity measures, a more diverse participant pool, and other computationally hard tasks such as the Travelling Salesperson Problem. More broadly, the framework employed here could be applied to other settings which may affect the cognitive resources available to a decision-maker, such as chronic stress, mental illness, and ageing. Finally, our results suggest that log predicted cortisol may be a promising trial-level measure for stress, suitable for experimental designs with more frequent cortisol sampling.

In conclusion, our study looked at the intersection of two crucial areas of decision-making research. Firstly, acute stress is known to be highly prevalent in our everyday lives and to significantly affect performance on a variety of decision-making tasks; however, the direction and size of these effects depends upon numerous moderating factors. Secondly, many decisions we have to make, including many important decisions, are computationally hard and often have enormous impacts on our well-being in areas such as finance and health. It is thus critical to understand how acute stress affects decision-making not only in easy cognitive tasks but also in computationally hard decisions. Our study addressed these gaps, providing empirical evidence that higher cortisol levels impaired decision performance in the KP, but that there was no interaction effect between acute stress and task complexity. While there is much work to be done to understand the generality of this result and the cognitive processes underpinning it, our study provides a first step towards a better understanding of how stress affects complex decisions.

Acknowledgements

We thank the anonymous reviewers for their time and helpful comments, Elizabeth Bowman for her support of the laboratory experiments, and our confederates for the TSST: Michelle Lee, Hassan Andrabi, Declan Grindrod, Elizabeth, and Chaoyi Chen. This was supported by the Australian Government Research Training Program Scholarship (Doroc) and the University of Melbourne Graduate Research Scholarship from the Faculty of Business and Economics (Doroc).

References

Allen, A. P., Kennedy, P. J., Cryan, J. F., Dinan, T. G., & Clarke, G. (2014, January). Biological and psychological markers of stress in humans: Focus on the Trier Social Stress Test. *Neuroscience & Biobehavioral Reviews*, *38*, 94–124. doi: 10.1016/j.neubiorev.2013.11.005

Arnsten, A. F. T. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience*, 13.

Arora, S., & Barak, B. (2009). *Computational Complexity: A Modern Approach*. Cambridge University Press.

Bali, A., & Jaggi, A. S. (2015, October). Clinical experimental stress studies: Methods and assessment. *Reviews in the Neurosciences*, *26*(5), 555–579. doi: 10.1515/revneuro-2015-0004

Cheeseman, P. C., Kanefsky, B., & Taylor, W. M. (1991). Where the Really Hard Problems Are. In *IJCAI*.

Chiu, S. K., Collier, C. P., Clark, A. F., & Wynn-Edwards, K. E. (2003, May). Salivary cortisol on ROCHE Elecsys immunoassay system: Pilot biological variation studies. *Clinical Biochemistry*, *36*(3), 211–214. doi: 10.1016/S0009-9120(02)00471-X

Duque, A., Cano-López, I., & Puig-Pérez, S. (2022, July). Effects of psychological stress and cortisol on decision making and modulating factors: A systematic review. *European Journal of Neuroscience*, *56*(2), 3889–3920. doi: 10.1111/ejn.15721

Franco, J. P., Doroc, K., Yadav, N., Bossaerts, P., & Murawski, C. (2022, July). Task-independent metrics of computational hardness predict human cognitive performance. *Scientific Reports*, *12*(1), 12914. doi: 10.1038/s41598-022-16565-w

Franco, J. P., Yadav, N., Bossaerts, P., & Murawski, C. (2021, September). Generic properties of a computational task predict human effort and performance. *Journal of Mathematical Psychology*, *104*, 102592. doi: 10.1016/j.jmp.2021.102592

Grace, C., Heinrichs, M., Koval, P., Gorelik, A., von Dawans, B., Terrett, G., … Labuschagne, I. (2022, November). Concordance in salivary cortisol and subjective anxiety to the trier social stress test in social anxiety disorder. *Biological Psychology*, *175*, 108444. doi: 10.1016/j.biopsycho.2022.108444

Het, S., Rohleder, N., Schoofs, D., Kirschbaum, C., & Wolf, O. (2009, August). Neuroendocrine and psychometric evaluation of a placebo version of the 'Trier Social Stress Test'. *Psychoneuroendocrinology*, *34*(7), 1075–1086. doi: 10.1016/j.psyneuen.2009.02.008

Johnson, S. B., Dariotis, J. K., & Wang, C. (2012, August). Adolescent Risk Taking Under Stressed and Nonstressed Conditions: Conservative, Calculating, and Impulsive Types. *Journal of Adolescent Health*, *51*(2), S34-S40. doi: 10.1016/j.jadohealth.2012.04.021

Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. (1993, February). The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*, *28*, 76–81. doi: 10.1159/000119004

Kudielka, B. M., Wüst, S., Kirschbaum, C., & Hellhammer, D. H. (2007). Trier Social Stress Test. In *Encyclopedia of Stress* (2nd ed., Vol. 3, pp. 776–781).

Labuschagne, I., Grace, C., Rendell, P., Terrett, G., & Heinrichs, M. (2019, December). An introductory guide to conducting the Trier Social Stress Test. *Neuroscience & Biobehavioral Reviews*, *107*, 686–695. doi: 10.1016/j.neubiorev.2019.09.032

Lai, G.-J., Yeh, Y.-c., Lin, C.-W., Hsu, W.-C., & Wu, J. (2017, May). The interactive influences of stress, modality of stimuli, and task difficulty on verbal versus visual working memory capacity. *Learning and Individual Differences*, *56*, 119–127. doi: 10.1016/j.lindif.2016.10.016

Lenow, J. K., Constantino, S. M., Daw, N. D., & Phelps, E. A. (2017, June). Chronic and Acute Stress Promote Overexploitation in Serial Decision Making. *The Journal of Neuroscience*, *37*(23), 5681–5689. doi: 10.1523/JNEUROSCI.3618-16.2017

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1. doi: 10.1017/S0140525X1900061X

Lighthall, N. R., Gorlick, M. A., Schoeke, A., Frank, M. J., & Mather, M. (2013, March). Stress modulates reinforcement learning in younger and older adults. *Psychology and Aging*, *28*(1), 35–46. doi: 10.1037/a0029823

Long, G. M., & Mo, S. S. (1971, February). Interaction of Arousal and Task Difficulty upon Perceptual Restriction. *Perceptual and Motor Skills*, *32*(1), 259–264. doi: 10.2466/pms.1971.32.1.259

Miller, R., Plessow, F., Kirschbaum, C., & Stalder, T. (2013, November). Classification Criteria for Distinguishing Cortisol Responders From Nonresponders to Psychosocial Stress: Evaluation of Salivary Cortisol Pulse Detection in Panel Designs. *Psychosomatic Medicine*, *75*(9), 832–840. doi: 10.1097/PSY.0000000000000002

Murawski, C., & Bossaerts, P. (2016, December). How Humans Solve Complex Problems: The Case of the Knapsack Problem. *Scientific Reports*, *6*(1), 34851. doi: 10.1038/srep34851

Nitsch, F. J., Sellitto, M., & Kalenscher, T. (2021, September). The effects of acute and chronic stress on choice consistency. *Psychoneuroendocrinology*, *131*, 105289. doi: 10.1016/j.psyneuen.2021.105289

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013, December). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, *110*(52), 20941–20946. doi: 10.1073/pnas.1312011110

Piccinini, G., & Scarantino, A. (2010, September). Computation vs. information processing: Why their difference matters to cognitive science. *Studies in History and Philosophy of Science Part A*, *41*(3), 237–246. doi: 10.1016/j.shpsa.2010.07.012

Plieger, T., & Reuter, M. (2020, September). Stress & executive functioning: A review considering moderating factors. *Neurobiology of Learning and Memory*, *173*, 107254. doi: 10.1016/j.nlm.2020.107254

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003, October). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology*, *28*(7), 916–931. doi: 10.1016/S0306-4530(02)00108-7

Richards, B. A., & Lillicrap, T. P. (2022, February). The Brain-Computer Metaphor Debate Is Useless: A Matter of Semantics. *Frontiers in Computer Science*, *4*, 810358. doi: 10.3389/fcomp.2022.810358

Rodrigues, S. M., LeDoux, J. E., & Sapolsky, R. M. (2009). The Influence of Stress Hormones on Fear Circuitry. *Annual Review of Neuroscience*, *32*(1), 289–313. doi: 10.1146/annurev.neuro.051508.135620

Sandi, C. (2013, May). Stress and cognition. *WIREs Cognitive Science*, *4*(3), 245–261. doi: 10.1002/wcs.1222

Sandner, M., Zeier, P., Lois, G., & Wessa, M. (2021, July). Cognitive emotion regulation withstands the stress test: An fMRI study on the effect of acute stress on distraction and reappraisal. *Neuropsychologia*, *157*, 107876. doi: 10.1016/j.neuropsychologia.2021.107876

Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How Do Glucocorticoids Influence Stress Responses? Integrating Permissive, Suppressive, Stimulatory, and Preparative Actions. *Endocrine Reviews*, *21*(1), 35.

Schoofs, D., Preuß, D., & Wolf, O. T. (2008, June). Psychosocial stress induces working memory impairments in an n-back paradigm. *Psychoneuroendocrinology*, *33*(5), 643–653. doi: 10.1016/j.psyneuen.2008.02.004

Schwabe, L., Wolf, O. T., & Oitzl, M. S. (2010, March). Memory formation under stress: Quantity and quality. *Neuroscience & Biobehavioral Reviews*, *34*(4), 584–591. doi: 10.1016/j.neubiorev.2009.11.015

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017, July). Toward a Rational and Mechanistic Account of Mental Effort. *Annual Review of Neuroscience*, *40*, 99–124.

Shields, G. S., Bonner, J. C., & Moons, W. G. (2015, August). Does cortisol influence core executive functions? A meta-analysis of acute cortisol administration effects on working memory, inhibition, and set-shifting. *Psychoneuroendocrinology*, *58*, 91–103. doi: 10.1016/j.psyneuen.2015.04.017

Shields, G. S., Trainor, B. C., Lam, J. C. W., & Yonelinas, A. P. (2016, September). Acute stress impairs cognitive flexibility in men, not women. *Stress*, *19*(5), 542–546. doi: 10.1080/10253890.2016.1192603

Starcke, K., & Brand, M. (2012, April). Decision making under stress: A selective review. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1228–1248. doi: 10.1016/j.neubiorev.2012.02.003

Taylor, S., Jaques, N., Weixuan Chen, Fedor, S., Sano, A., & Picard, R. (2015, August). Automatic identification of artifacts in electrodermal activity data. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1934–1937). Milan: IEEE. doi: 10.1109/EMBC.2015.7318762

Vecchio, R. P. (1990). Theoretical and Empirical Examination of Cognitive Resource Theory. *Journal of Applied Psychology*, *75*(2), 141–147.

Vogeser, M., Kratzsch, J., Ju Bae, Y., Bruegel, M., Ceglarek, U., Fiers, T., ... Ostlund, R. E. (2017, June). Multicenter performance evaluation of a second generation cortisol assay. *Clinical Chemistry and Laboratory Medicine (CCLM)*, *55*(6), 826–835. doi: 10.1515/cclm-2016-0400

Watson, D., Anna, L., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. doi: 10.1037/0022-3514.54.6.1063

Wemm, S. E., & Wulfert, E. (2017, March). Effects of Acute Stress on Decision Making. *Applied Psychophysiology and Biofeedback*, *42*(1), 1–12. doi: 10.1007/s10484-016-9347-8

Yadav, N., Murawski, C., Sardina, S., & Bossaerts, P. (2018, June). *Phase transition in the knapsack problem* (No. arXiv:1806.10244). arXiv.

Yerkes, R. M., & Dodson, J. D. (1908, November). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482. doi: 10.1002/cne.920180503

Yu, R. (2016, June). Stress potentiates decision biases: A stress induced deliberation-to-intuition (SIDI) model. *Neurobiology of Stress*, *3*, 83–95. doi: 10.1016/j.ynstr.2015.12.006