



# Stress or failure? An experimental protocol to distinguish the environmental determinants of decision-making

Martina Vecchi<sup>1</sup> · Nicolai Vitt<sup>2</sup>

Received: 30 September 2023 / Revised: 18 March 2024 / Accepted: 14 May 2024 /  
Published online: 15 July 2024  
© The Author(s) 2024

## Abstract

Are economic decisions affected by short-term stress, failure, or both? Such effects have not been clearly distinguished in previous experimental research, and have the potential to worsen economic outcomes, especially in disadvantaged socioeconomic groups. We validate a novel experimental protocol to examine the individual and combined influences of stress, failure, and success. The protocol employs a 2×3 experimental design in two sessions and can be used online or in laboratory studies to analyse the impact of these factors on decision-making and behaviour. The stress protocol was perceived as significantly more stressful than a control task, and it induced a sizeable and significant rise in state anxiety. The provision of negative feedback (“failure”) significantly lowered participants’ assessment of their performance, induced feelings of failure, and raised state anxiety.

**Keywords** Acute stress · Failure · Online experiment · Experimental protocol

**JEL Classification** C9 · D91

---

✉ Martina Vecchi  
m.vecchi@soton.ac.uk

Nicolai Vitt  
nicolai.vitt@bristol.ac.uk

<sup>1</sup> Department of Economics, University of Southampton, Southampton, UK

<sup>2</sup> School of Economics, University of Bristol, Bristol, UK

## 1 Introduction

Individuals making economic decisions often operate in stressful environments and with previous experiences of both success and failure on their minds. The experience of situational acute stress can create anxiety, divert attention, and impact the body and mind.<sup>1</sup> Additionally, individuals who recall past poor decisions may experience reduced confidence and diminished willingness to take further actions. Prior failures and acute stress can have complex simultaneous effects on individuals' decisions. Previous experimental studies have been unable to clearly distinguish between the influences of these factors. We conducted a behavioural online experiment to validate a novel experiment protocol that separates the effects of short-term stress and failure, allowing researchers to assess their effects individually and in combination.

Acute stress can arise from the decision-making process itself and from external factors such as financial and relationship concerns, and high-pressure work environments. Stress is the body's response to a short-term demand or pressure and manifests as physiological changes such as a rapid heart rate and as psychological effects such as feelings of anxiety (Daviu et al., 2019; Giannakakis et al., 2017). Experimental studies in behavioural economics and the social and biological sciences have shown that acute stress affects decision-making in settings from financial to health care (Bendahan et al., 2017; Cahlíková & Cingl, 2017; Delaney et al., 2014; Haushofer et al., 2018; Rutters et al., 2009; Von Dawans et al., 2012). Several found that stress caused a shift in cognitive effort towards the stressor and away from other tasks (Allen & Armstrong, 2006; Cohen, 1980; Starcke et al., 2016), and that stress impaired working memory and cognitive flexibility (see Shields et al., 2016, for a review). Self-preservation instincts can also translate to a greater effort to preserve financial and other resources when stressed (Dickerson & Kemeny, 2004).

Failures are past experiences that we view as unsuccessful because we did not meet our own or others' expectations. Feelings of failure may arise from past decisions that did not yield favorable outcomes or from a broader perception of inadequacy, possibly due to unemployment, stalled career progress, or financial challenges. Experimental research has shown failure to affect how we make decisions (Buser, 2016; Buser & Dreber, 2016; Cassar & Klein, 2019; Gill & Prowse, 2014). Failures have been found to provoke negative emotions such as guilt and shame (Bohns & Flynn, 2013; Carver & Scheier, 1990), alter individuals' sense of self-worth (Crocker & Park, 2004; Crocker & Wolfe, 2001; Heatherton & Polivy, 1991), and drive them to abandon the task (Crocker & Wolfe, 2001), or attempt to compensate for prior losses (Thaler & Johnson, 1990). A history of failure may generate a self-reinforcing cycle of "poor" choices and setbacks, contributing to the persistence of poverty (Stevens, 1999) and creating a barrier to socioeconomic mobility (Corak, 2013). Prior experience of failure or success may influence whether a stressor is perceived as a threat or as a challenge. Moreover, stress may impact the processing of feedback, potentially reducing the degree to which participants learn, particularly

<sup>1</sup> In the following, acute stress, short-term stress, and stress will be used interchangeably when describing stress or stressors.

from negative feedback (Petzold et al., 2010; Porcelli & Delgado, 2017; Raio et al., 2017).

Previous experimental research could not separate the effects of acute stress from the effects of failure because the protocols incorporated elements of both. The Trier Social Stress Test (Kirschbaum et al., 1993), commonly used to induce stress, asks participants to give a presentation and complete a mental arithmetic task and informs the participants that their performance will be evaluated, likely provoking fear of failure. Other stress protocols have asked participants to complete unsolvable riddles or extremely difficult cognitive tasks (Habhab et al., 2009; Rutters et al., 2009) in which failure is used as the stressor. Similarly, experimental studies on failure have relied on competitive settings, which inherently entail elements of stress (Buckert et al., 2017) and thus are unable to clearly distinguish the effects of failure from those of stress.

Stress and failure are not intrinsically linked. Within educational and occupational environments, elevated workloads can elicit stress independently of any failure. Similarly, perceptions of failure, such as those stemming from insufficient stimulation or negative performance feedback in academic or professional settings, do not invariably coincide with stress. Given that stress and failure can exert distinct influences on decision-making processes, a more effective protocol to discern their respective roles will help pinpoint affected groups and contexts, informing targeted policy interventions.

Our experiment employs a  $2 \times 3$  two-session design to manipulate participants' decision-making environment. In the first treatment, we manipulate whether participants engage in a stress-inducing task during either the first or second session, thereby introducing exogenous variation in acute stress levels. The stress task simulated common stressors faced by students, resembling assessment test questions for job or graduate school applications. During the 10-min task participants answered various cognitive questions, facing financial incentives with potential losses, cognitive and time pressures, and distractions. In the second treatment, we manipulate whether participants received no performance feedback, feedback relative to a low threshold (success), or relative to a high threshold (failure) before the decision-making tasks in the second session. This feedback is not intended to induce stress, but rather to introduce variation in perceived success or failure.

The two-session design is crucial for inducing failure/success separately from stress. All participants received feedback related to the stress task, ensuring consistency. Those who completed the stress task in the first session received success or failure feedback in the second session, when no longer exposed to stress, enabling us to isolate the individual effects of failure and success. Conversely, participants completing the stress task in the second session received feedback while still experiencing stress. This design generates six different decision environments and allows between-subject and within-subject comparisons across the two sessions.

We find strong evidence that the protocol induced short-term psychological stress among the participants, who reported significantly greater levels of stress from the stress-inducing task than from the control task and reported statistically significant increases in their anxiety following the stress task. We also find strong evidence that negative feedback affected participants' assessments of their performance, triggered

emotional responses, and induced stronger feelings of failure. The response to positive feedback, on the other hand, did not differ strongly from the response to no feedback.

Our study makes several contributions to the literature. First, we develop an experimental design that allows researchers to distinguish the individual effects of acute stress, failure, and success. We also establish a protocol for experiments that induces realistic cognitive stress in student populations and propose a novel feedback protocol that allows researchers to vary participants' perceptions of success and failure and induce emotional responses. These protocols can be used in online and laboratory settings to investigate the impacts of stress and failure separately and in combination.

This paper is structured as follows. Section 2 describes the experimental design. Section 3 presents the results. Section 4 concludes.

## 2 Experimental design

### 2.1 Sample and recruitment

Our online experiment validated a protocol to test the effects of stress and feedback on decisions participants made in two sessions that were scheduled at least seven days apart.<sup>2</sup> The first session lasted approximately 50 min and the second session lasted approximately 60 min. We conducted the experiment between 04 June and 31 July, 2020, and recruited students at Pennsylvania State University after ethical approval of the study from the university's Institutional Review Board. We pre-registered the experiment and hypotheses tested in the AEA RCT registry under the following trial ID: AEARCTR-0005946. Details can be found at <http://www.socialscineregistry.org/trials/5946>. The experiment design was pre-tested in May 2020 with a small sample of students.

In the experiment, participants performed four different decision-making outcome tasks. Difficulties associated with the COVID-19 pandemic restricted our ability to recruit participants in 2020, leading to a smaller sample size (final sample of 269 students) than specified and reducing the power of the study to detect treatment effects on the decision-making outcome tasks. Future research will focus on the outcomes of decisions under stress and failure/success after collecting data from additional participants. It is essential to note the heightened background stress during the pandemic period, thus the potential for additional increase in stress during the experiment may have been limited. Both the stress and control groups in our study operated within this elevated stress context. Consequently, the impact of our stressor

<sup>2</sup> This validation study was planned as a lab experiment but was adapted to an online setting due to the COVID-19 pandemic. Certain elements of the initial experimental plan could not be incorporated in the online setting. In particular, the measurement of participants' heart rate to capture the physiological response to acute stress was not possible in the online experiment.

may be considered a conservative estimate, representing a lower bound of its potential effect in scenarios characterized by lower background stress.

## 2.2 Procedure and randomization

As shown in Table 1, our  $2 \times 3$  two-session experiment design produced six conditions that varied in the timing of stress and experience of failure/success. The experimental conditions were randomized at the participant level and introduced random variation in the decision environment of participants.

Treatment 1 varied the timing of the stress task. At the beginning of the first experiment session, half of the participants were assigned the stress-neutral control task (S2 in Table 1): answer demographics questions, read a short text and answer questions about it. The other half were assigned the stress task and completed an incentivized cognitive task (S1). In the second session about a week later, participants were assigned to the task they did not complete in the first session, so each participant completed one control session and one stress session.

To induce feelings of failure and success, Treatment 2 introduced feedback regarding participants' performance in the stress task. The incentive structure for the stress task involved a potential deduction from participants' payoff if they performed below a predetermined threshold. Treatment 2 varied whether participants received feedback about their performance in the stress task relative to the threshold and the difficulty of the threshold. Feedback was presented in the second session. At that point, half the participants had completed the stress task that day, and the other half had completed the stress task approximately one week earlier during the first session. This temporal break between stress and feedback allowed us to analyse the impacts of stress and failure/success separately. The three feedback conditions were no feedback (S1-N, S2-N in Table 1), a success condition that provided feedback relative to a low threshold for success (S1-S, S2-S), and a failure condition that gave feedback relative to a high threshold for success (S1-F, S2-F). These distinct feedback conditions were instrumental in inducing feelings of failure and success, enabling us to isolate their effects. The experiment presented two experimental conditions during the first session (stress and no-stress per Treatment 1), and six in the second session (stress and no-stress combined with the three types of feedback from Treatment 2). In the second stage of each session, participants completed the decision-making outcome tasks. At three points in each session (at the beginning, after completion of the task, and at the end), participants completed the six-item short-form state anxiety inventory developed by Marteau and Bekker (1992).

They also completed brief questionnaires. First-session questions addressed their perceptions of the assigned task. Second-session questions again addressed their perceptions of the task assigned that day and collected information about their demographic characteristics, actual stress experiences, failures, and decision-making processes.

The full timeline of the experimental sessions is displayed in Online Appendix A. The experimental instructions shown to participants can be found in Online Appendix B.

Table 1 Experimental treatments and conditions  
(a) During the first session:

Treatment 1: Timing of stress task	
Stress in the 1st session (S1)	Stress in the 2nd session (S2)
S1: Stress only	S2: Control

(b) During the second session:

Treatment 1: Timing of stress task	
Stress in the 1st session (S1)	Stress in the 2nd session (S2)
S1-N: Control	S2-N: Stress only
S1-S: Success only	S2-S: Stress & Success
S1-F: Failure only	S2-F: Stress & Failure

Treatment 2: Feedback	No feedback (N)
	Low threshold (S)
	High threshold (F)

Participants received a fixed participation fee of \$5 at the end of their second session. 50% of the participants received payment based on their performance in the stress task. The other 50% was paid based on one of the decision-making outcome tasks. Minimum compensation was \$5, maximum compensation was \$45, and the mean for the study was \$20.96. All payments were made via PayPal within 48 hours of completion of the second session. This arrangement minimized potential attrition effects between the first and second session, avoided influence from wealth and income effects, and provided no information about which task would be for payment until the end of the second session.

### 2.3 Stress protocol

The stress protocol was designed to mimic stressors commonly experienced by students and was framed as a “block of several tasks that are similar to test questions you may face in assessment tests when applying for jobs or for graduate school.”

To induce stress, we incorporated a financial incentive with potential losses of payoff, cognitive pressure, time pressure, and distractions in the stress task. All have been shown to induce stress in study participants. The stress task required them to complete up to 18 short cognitive tasks in 10 min, and they were penalized for incorrect answers and leaving tasks undone. As a distraction, the program intermittently displayed brief incentivized knowledge and arithmetic questions on the screen during the ten-minute task block to induce additional stress.

The task was loosely based on the stress task used in Vitt et al. (2021), which was perceived by the study sample of low-income mothers as significantly more stressful than a control task and induced a significant increase in participants' heart rate. We adapted this protocol to the different population of interest, university students. Differently from other studies (e.g. the Trier Social Stress Test by Kirschbaum et al., 1993, or the cold pressor test used in Delaney et al., 2014), and similarly to Vitt et al. (2021), our aim is not to trigger strong physiological responses (such as the stress hormone cortisol), but to mimic stressors university students experience in real life.

Participants were initially allocated a maximum potential incentive of 5000 tokens (\$40). Participants' performance in the cognitive tasks and pop-up questions determined how much of the initial endowment they “lost.” The incentives are framed as a loss to avoid inducing positive feelings from “winning.” Participants could lose up to 4050 tokens (\$32.40) from the 18 cognitive tasks and 200 tokens (\$1.60) from the pop-up questions. In addition, 750 tokens (\$6.00) would be deducted if their performance in the stress task was below an undisclosed threshold. The level of this threshold was experimentally varied by Treatment 2 (described below) to induce failure or success. Participants were informed that they would be randomly assigned to a group of participants who all faced the same threshold.

A detailed description of the stress task and the incentive structure, as well as several sample screenshots can be found in Online Appendix C.

## 2.4 Control tasks

In the control sessions, participants were given 4 min to complete a short demographics questionnaire. They then had 10 min to read a text about the possibility of life on Mars (518 words) and a text about the evolution of languages (629 words) and to answer three simple questions about each text.

This task required participants to pay attention but was not meant to induce stress. The time required was equal to the time required of participants to review the instructions for the stress task plus the time allotted to complete it. No financial or time pressure was induced by the control task.

## 2.5 Feedback on failure or success

Treatment 2 assigned each participant in the second session to a feedback condition. Following the stress or control task, the program informed participants assigned to the no-feedback control group (S1-N and S2-N) that there would be a 60-second wait before proceeding. For participants assigned to the feedback treatments (S1-S, S2-S, S1-F, or S2-F), the screen displayed failure/success feedback as a statement that noted only whether they had met the threshold to avoid the 750-token penalty. It was displayed for 60 seconds. Screenshots are presented in Figures C.5 - C.7 of Online Appendix C.

The threshold in the success condition (S1-S, S2-S) was 1,250 tokens such that approximately 95% of participants were expected to succeed. The threshold in the failure condition (S1-F, S2-F) was 4300 tokens such that approximately 95% of participants were expected to fail. This random variation in perception of success and failure (conditional on participants' performance) was independent from the timing of the stress task, allowing us to overcome any confounding factors like their ability, and to identify the influence of success/failure.

We assess our protocol's effectiveness in manipulating feelings of failure by collecting participants' self-reports of their perceived successfulness, performance, and post-feedback emotions. We expected failure feedback to increase feelings of failure, decrease self-assessments of performance, and induce negative emotions. Details on these measures of failure / success are in Online Appendix D.2.

## 2.6 Empirical approach

To be able to identify the impacts of stress and failure individually and when combined, participants completed the outcome tasks of interest at the end of the first and the second sessions. This design allows a between-subject and within-subject comparison across the two sessions. The fixed-effects model used for this analysis and the corresponding power calculations are presented in Online Appendices E.1 and F. Furthermore, to deal with non-random attrition between sessions, we used a



Heckman model of selection (Heckman, 1979; Wooldridge, 2010), which is outlined in Online Appendix E.2.

### 3 Results

#### 3.1 Descriptive statistics

The demographic characteristics of the sample in our validation study are provided in Table G.1 in Online Appendix G.1. The average age of participants was 23.5 years with 66% of participants identifying as female and 32% as male. All participants were students; 72% were undergraduates and 28% were post-graduates. Our comparison of responses from participants assigned to the stress condition in the first session (S1) with those assigned to the stress condition in the second session (S2) shows a significant difference in average participant age. We report our demographic comparisons by feedback condition in Table G.2 in Online Appendix G.1.

Of the 317 participants who completed the first session, 269 (84.86%) returned and successfully completed the second session, 12 (3.79%) provided invalid or incomplete responses in the second session and were dropped, and 36 (11.36%) did not return for the second session. In Online Appendices G.2 and G.3 we further examine these attrition issues and show that our results are robust to accounting for non-random attrition.

#### 3.2 Effectiveness of stress protocol

Our protocol was designed to induce acute stress, and we used psychological stress measures to analyse its effectiveness. As shown in Table 2, the stress task was rated as substantially more stressful than the control task on a Likert scale from 1 (not at all) to 5 (very much). The mean stressfulness score was 3.5 for the stress task and 2.1 for the control task. The sizeable difference of 1.4 is equivalent to 1.33 standard deviations and is highly statistically significant. We found only small and not statistically significant differences in perceived stress by feedback condition. The results in Table 3 show that the stress task was perceived, on average, as significantly less relaxing, less easy, more difficult, less enjoyable, less successful, and more tiring than the control task.

Figure 1 presents mean state anxiety scores in the stress and control conditions for the three measurements taken in each session. A difference-in-difference comparison of state anxiety levels is provided in Table G.3 of Online Appendix G. We find baseline mean anxiety levels (T1) of 37.6 and 36.8 in the stress and control conditions respectively, which are not statistically different and are in line with values reported in Marteau and Bekker (1992). The second anxiety measurement (T2) was conducted shortly after completion of the stress or control task. At that point, the mean anxiety level of the control participants was 36.7, nearly unchanged from the baseline level. The mean anxiety level of the stress-task participants rose from 37.6 to 46.7, a substantial increase of 9.1 (0.77 baseline

**Table 2** Perceived stressfulness of the stress/control task

	(1)	(2)	(3)	Differences:	
	No feedback	Success	Failure	(2) – (1)	(3) – (1)
Stress	3.483 (0.082)	3.583 (0.178)	3.763 (0.169)	0.100 (0.196)	0.280 (0.187)
Control	2.181 (0.082)	1.902 (0.151)	2.186 (0.153)	–0.278 (0.171)	0.005 (0.173)
Difference:	1.302***	1.681***	1.577***	0.379	0.275
Stress – Control	(0.113)	(0.234)	(0.228)	(0.272)	(0.257)
Observations	342	77	81		
Participants	250	77	81		

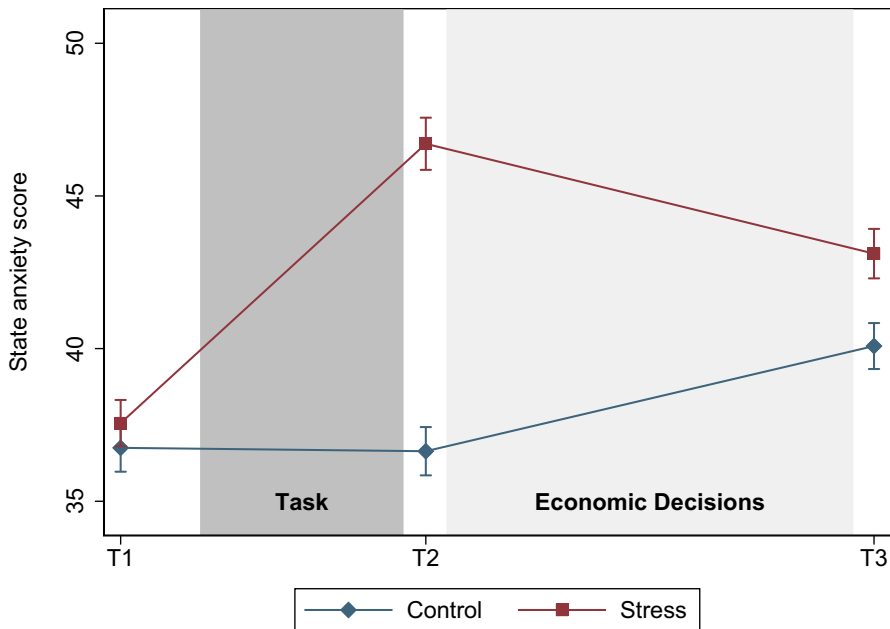
Note: Means and mean differences were obtained using the sample of participants who responded to the task stressfulness question in both sessions. Participants with a performance level above the high or below the low threshold were excluded from the sample for simplicity. Standard errors were clustered at the participant level and are shown in parentheses. Significance levels are indicated as follows: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

standard deviations). We find that the increase in anxiety among stress-task participants and the difference in anxiety levels between the control and stress-task participants are both statistically significant. They are larger than ones reported for the Trier Social Stress Test for Groups (increase by approximately 7 points, Von Dawans et al., 2011), and for stress protocols based on unsolvable arithmetic

**Table 3** Perceptions of the stress/control task

	Stress	Control	Difference	Observations	Participants
Stressful	3.540 (0.069)	2.136 (0.066)	1.404*** (0.089)	500	250
Relaxing	2.235 (0.067)	3.195 (0.064)	–0.960*** (0.082)	502	251
Easy	2.779 (0.070)	4.237 (0.055)	–1.458*** (0.081)	498	249
Difficult	3.248 (0.070)	1.776 (0.052)	1.472*** (0.085)	500	250
Enjoyable	2.520 (0.070)	3.064 (0.066)	–0.544*** (0.083)	500	250
Successful	2.867 (0.061)	3.791 (0.064)	–0.924*** (0.080)	498	249
Tiring	2.908 (0.074)	2.156 (0.068)	0.752*** (0.084)	500	250

Note: Perceptions of the stress / control task were scored from 1 for ‘not at all’ to 5 for ‘very much’. Means and mean differences were obtained using the sample of participants who responded to the task perception question in both sessions. Participants with a performance level above the high or below the low threshold were excluded from the sample for simplicity. Standard errors were clustered at the participant level and are shown in parentheses. Significance levels are indicated as follows: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$



**Fig. 1** State anxiety response to the stress/control task. Note: Means were calculated for the state anxiety scores at the three measurement points. Bands indicate  $\pm 1$  standard error. Participants with a performance level above the high or below the low threshold were excluded here for simplicity

tasks (increase by approximately 3 points, Rutters et al., 2009). By the end of the session (T3) the gap in anxiety levels between the stress and control conditions closed somewhat, indicating that the acute stress dissipated slightly when participants' attention was shifted from the task.

Overall, we find strong evidence that the stress protocol was effective in inducing mild short-term psychological stress. To contextualize the effect sizes observed in our study, Tables H.1 and H.2 in Appendix H compare the impact of the stress protocol in our study on perceived stress and state anxiety with findings for other experimental stressors from previous laboratory investigations and randomized controlled trials. Our study reveals that our stress protocol induces a notable increase in perceived stress, equivalent to 1.35 control group standard deviations. In contrast, previous studies utilizing the Trier Social Stress Test report comparatively lower effects on perceived stress, ranging from no effect to 0.3 standard deviations. Studies employing economic games or presenting riddles as stressors demonstrate effects on perceived stress within the range of 0.33 to 4.6 control standard deviations. Furthermore, we observe a significant elevation in state anxiety, as measured by the short-form State-Trait Anxiety Inventory (STAI), with a difference of 10.07 points between the stress and control group following the task. This increase surpasses those typically reported in other studies, which generally range from 2.6 to 4.7 points difference from the control group, with one exception being a study on anticipatory stress that found a 12.15 point difference from the control group.

Previous research suggests the presence of gender differences in the response to stress (e.g., Taylor et al., 2000). Thus, we explore in Tables G.16, G.17, G.18 and Figure G.6 of Online Appendix G.4 whether the effectiveness of the stress protocol differs between male and female participants. We find the psychological stress response to the stress task to not differ significantly between men and women.

While the present study did not measure physiological stress due to the online setting in which the experiment was conducted, a lab experiment conducted by the authors in 2023 (manuscript in preparation) provides evidence on the physiological stress response to the protocol. As part of the experiment, we measured participants' heart rate and observed a significantly increased heart rate during the stress task (compared to the control task). Thus, the stress protocol induces both psychological and physiological stress.

### 3.3 Effectiveness of feedback

By randomizing the feedback and threshold for success we aim to induce feelings of failure or success among participants. We analyse whether the feedback provided novel information to participants about their performance and whether it modified perceived success in the stress task.

For the feedback to provide novel information, the protocol must prevent participants from keeping track of their performance. We find that participants' two self-assessment ratings are only weakly correlated with actual performance: a correlation coefficient of 0.068 for the Likert-scale measure and  $-0.120$  for expectations about losing tokens. Figures G.1 and G.2 in Online Appendix G show the distribution of participants' performance prior to deductions. Given the mean performance score of 2386 tokens, participants lost, on average, 61.5% of the 4250 tokens associated directly with performance. With a standard deviation of 621 tokens (14.6%), the distribution is sufficiently narrow for feedback to provide meaningful information to participants. Nearly all of the participants (97.8%) performed at levels between the low and high thresholds. Table G.4 in Online Appendix G summarizes participants' self-assessments of likely exceeding the success threshold. In the no-feedback condition, 66% (70%) of those who exceeded (did not exceed) their assigned thresholds expected to exceed the threshold. There is no significant difference, indicating that success was difficult to predict without feedback.

Table 4 compares perceived success in the stress and control tasks by feedback condition rated on a Likert scale of 1 (not at all successful) to 5 (very much). Participants in the stress condition who received failure feedback reported the lowest success scores: a mean of 2.4 relative to 2.9 for reports by participants who received no feedback, a statistically significant difference (0.56 standard deviations). Though the feedback did not relate to the control task, S1 participants who received failure feedback rated their success lower than S1 participants who received no feedback. The difference of 0.3 is smaller and is not statistically significant. Success feedback instead did not lead to significantly greater self-perceptions of success.

At the end of the second session, participants were asked to judge their performance in the stress task using a Likert scale of 0 (very bad) to 4 (very good). Table 5

**Table 4** Perceived successfulness of the stress/control task

	(1)	(2)	(3)	Differences:	
	No feedback	Success	Failure	(2) – (1)	(3) – (1)
Stress	2.932 (0.071)	3.028 (0.145)	2.405 (0.161)	0.096 (0.161)	–0.526** (0.175)
Control	3.830 (0.079)	3.927 (0.145)	3.512 (0.149)	0.097 (0.164)	–0.319 (0.168)
Difference:	–0.898***	–0.899***	–1.106***	–0.001	–0.208
Stress – Control	(0.096)	(0.205)	(0.219)	(0.231)	(0.241)
Observations	341	77	80		
Participants	249	77	80		

Note: Means and mean differences were obtained using the sample of participants who responded to the task successfulness question in both sessions. Participants with a performance level above the high or below the low threshold were excluded from the sample for simplicity. Standard errors were clustered at the participant level and are shown in parentheses. Significance levels are indicated as follows: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

**Table 5** Perceived performance in the stress task

	(1)	(2)	(3)	Differences:		
	No feedback	Success	Failure	(2) – (1)	(3) – (1)	(3) – (2)
S2: Stress in Session 2	1.979 (0.157)	2.132 (0.161)	1.366 (0.120)	0.153 (0.224)	–0.613** (0.197)	–0.766*** (0.200)
S1: Stress in Session 1	2.574 (0.145)	2.767 (0.166)	2.295 (0.161)	0.193 (0.220)	–0.279 (0.217)	–0.472* (0.231)
Difference:	–0.596**	–0.636**	–0.930***	–0.040	–0.334	–0.294
Session 2 – Session 1	(0.213)	(0.231)	(0.201)	(0.314)	(0.293)	(0.306)
N	94	81	85			

Note: Performance perceptions were scored from 0 for ‘very bad’ to 4 for ‘very good’. Participants with a performance level above the high or below the low threshold were excluded from the sample for simplicity. Standard errors are shown in parentheses. Significance levels are indicated as follows: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

presents a comparison of their responses by feedback condition. S2 participants who received failure feedback rated their performance significantly lower than S2 participants who received no feedback (difference of 0.6) and positive feedback (difference of 0.8 points/0.76 SDs). S1 participants rated their performance significantly higher than the S2 participants, suggesting that participants’ self-assessments improve with time. S1 participants who received failure feedback rated their performance lower on average than S1 participants who received no feedback (difference of 0.3) and positive feedback (difference of 0.5 points/0.45 SDs), and only the latter difference was statistically significant. We also once again find that responses following positive feedback were not significantly different from responses following no feedback.

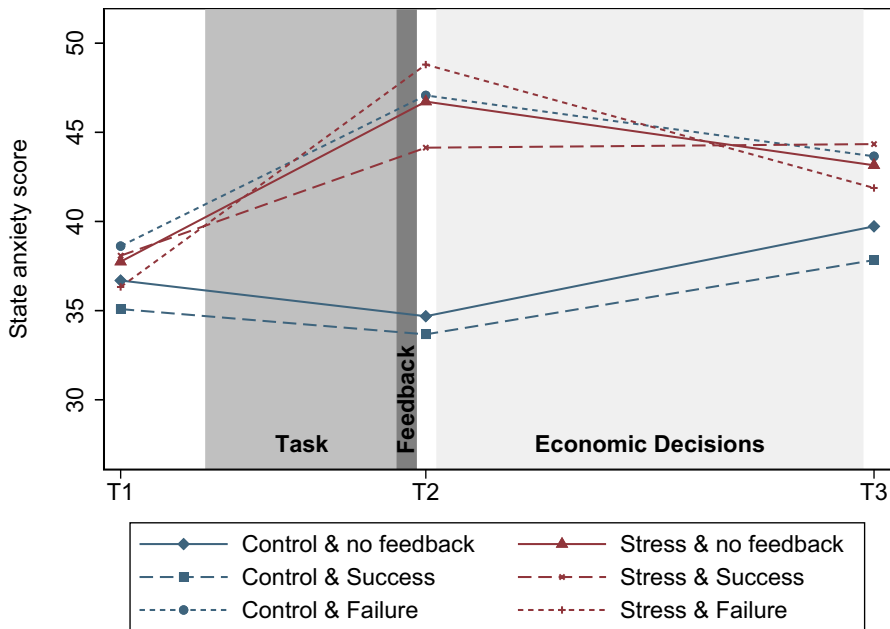
**Table 6** Recalled emotions after receiving feedback

	(1) Success	(2) Failure	Difference: Failure – Suc- cess
Pleased	3.580 (0.114)	1.894 (0.096)	– 1.686*** (0.149)
Angry	1.753 (0.105)	3.106 (0.125)	1.353*** (0.163)
Calm	3.321 (0.115)	2.447 (0.110)	– 0.874*** (0.159)
Anxious	2.395 (0.143)	2.906 (0.121)	0.511** (0.188)
Confident	3.185 (0.118)	2.000 (0.093)	– 1.185*** (0.150)
Disappointed	2.123 (0.125)	3.929 (0.106)	1.806*** (0.163)
Encouraged	3.284 (0.122)	1.929 (0.091)	– 1.355*** (0.152)
Sad	1.630 (0.102)	2.918 (0.122)	1.288*** (0.159)
Embarrassed	1.704 (0.113)	3.047 (0.133)	1.343*** (0.174)
Successful	3.457 (0.105)	1.812 (0.088)	– 1.645*** (0.137)
N	81	85	

Note: Emotions after receiving the feedback were scored from 1 for 'not at all' to 5 for 'very much'. Participants with a performance level above the high or below the low threshold were excluded from the sample for simplicity. Standard errors are shown in parentheses. Significance levels are indicated as follows: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

Table G.5 in Online Appendix G compares participant responses regarding their expected token losses by feedback condition. Among S2 participants, those who received failure feedback rated their expected loss of tokens significantly higher than those who received success feedback.

Finally, as shown in Table 6, we find that participants who received failure feedback felt significantly less pleased, calm, confident, encouraged, and successful and felt more angry, anxious, disappointed, sad, and embarrassed than participants who received success feedback. We further observed a significant increase from baseline in the state anxiety of participants who received failure feedback (see Fig. 2). The participants in the control condition who received failure feedback had a greater average increase in state anxiety than participants in the stress condition who received success feedback or no feedback. These results confirm the ability of our feedback protocol to induce negative emotions and that success feedback does not mitigate increases in anxiety triggered by stress.



**Fig. 2** State anxiety response to the stress/control task and the feedback. Note: Means were calculated for the state anxiety scores at the three measurement points. Participants with a performance level above the high or below the low threshold were excluded from the sample for simplicity

Overall, these findings provide strong evidence that the provision of feedback affected participants' assessment of their performance, their perceptions of failure/success, and their emotions. Failure feedback, in particular, induced a strong response, while the response to success feedback did not differ substantially from receiving no feedback. Previous experimental evidence highlights gender-specific responses to successes and failures (Buser, 2016). However, in Tables G.19, G.20, G.21 and Figure G.7 of Online Appendix G.4 we find no substantial differences in the response to the feedback between male and female participants.

## 4 Conclusion

In this study, we develop an experimental protocol to identify the individual and combined effects of acute stress and failure on decision-making. The protocol uses a two-session design to vary participants' exposure to stress, failure, and success. Exogenous variation in acute stress levels is introduced by assigning an incentivized cognitive task designed to induce mild stress. Half of the participants complete the stress task in the first session and half complete it in the second session. Those in a session not assigned to the stress task complete a stress-neutral control task during the same time period.

The experiment introduces variation in participants' perceptions of success or failure via a feedback protocol in the second session. The incentive for the stress

task deducts a portion of the performance payoff for participants whose performance falls below an undisclosed threshold. The threshold level and the provision of feedback are randomly assigned. Participants receive either no feedback, feedback with a low threshold level (success condition), or feedback with a high threshold level (failure condition).

We validate the protocol using an online experiment with a sample of 269 university students and find that participants perceived the incentivized stress task as significantly more stressful than the control task. State anxiety also increased substantially and significantly after completing the stress task. These results provide strong evidence that our stress protocol induced short-term psychological stress.

Participants who received failure feedback reported significantly lower self-assessments of their performance and success in the cognitive stress task, significantly greater anxiety, and strong emotional responses. Responses from participants who received success feedback did not differ substantially from responses of participants who received no feedback. Thus, we find strong evidence that our provision of negative feedback negatively affected participants' perceptions of themselves and evoked negative emotions.

Due to its online setting our validation experiment was unable to study the physiological stress response. The most common measure of physiological stress, the hormone cortisol, is likely not suitable in our context. Previous research has shown that not all psychological stressors produce a cortisol response (Dickerson & Kemeny, 2004), and any cortisol response to our stress protocol would likely be small and difficult to detect, as the protocol aims to induce mild levels of stress similar to those frequently faced in everyday life. Previous research suggests that participants' heart rate may be more suited as a physiological measure of mild stress (Vitt et al., 2021). In a lab experiment conducted in 2023 (manuscript in preparation) we use the stress protocol presented here, and observe a significant increase in heart rate during the stress task compared to the control task.

The ability of our protocol to induce short-term stress and feelings of failure among participants in online and laboratory experiments will allow future studies of the impacts of those two dimensions on decision-making. Our findings indicate that receiving success feedback does not alter individuals' assessments of their performance or how successful they feel. Therefore, it would be more efficient to concentrate future studies solely on a comparison between failure feedback and no feedback.

Stress and failure likely affect decision-making differentially in a variety of settings. At school and at work, stress could improve individuals' focus on a task, increase their cognitive load, and decrease their attention on other tasks. Such stress could be addressed by adapting the workloads and providing coping strategies. Failure, on the other hand, likely affects self-worth and self-esteem, decreases confidence, and induces compensatory behaviours. Redesigning goal settings and feedback structures could improve individuals' reactions to a poor performance. Future studies could use the protocol in this study to examine the relative importance of stress and failure for various types of decisions, thereby uncovering effective interventions.



**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40881-024-00172-8>.

**Acknowledgements** We thank the LEMA lab at Penn State University, and in particular Anthony Kwasnica and Rashmi Sharma, for their help with the study. We thank Michèle Belot, Pablo Brañas-Garza, Steven Dieterle, Lorenz Götte, Edward Jaenicke, Jonathan James, Tatiana Kornienko, Trent Smith, Andreas Steinhauer, Till Stowasser, Roberto A. Weber, Conny Wollbrant, Katherine Y. Zipp, as well as seminar participants at the SGPE conference, the University of Stirling, the University of Edinburgh, the Pennsylvania State University and the University of Bath for their helpful comments.

**Funding** We gratefully acknowledge funding of this project from the University of Edinburgh, School of Economics and from Pennsylvania State University, Department of Agricultural Economics, Sociology and Education. Nicolai Vitt furthermore received funding from the Economic and Social Research Council, award 1651922. This work was furthermore supported by the USDA National Institute of Food and Agriculture and Hatch Appropriations under Project number PEN04709 and Accession number 1019915.

**Data availability** The replication material for the study is available at: <https://doi.org/10.17605/OSF.IO/D8TAE>.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Allen, T. D., & Armstrong, J. (2006). Further examination of the link between work-family conflict and physical health. *American Behavioral Scientist*, 49(9), 1204–1221.
- Bendahan, S., Goette, L., Thoresen, J., Loued-Khenissi, L., Hollis, F., & Sandi, C. (2017). Acute stress alters individual risk taking in a time-dependent manner and leads to anti-social risk. *European Journal of Neuroscience*, 45(7), 877–885.
- Bohns, V. K., & Flynn, F. J. (2013). Guilt by Design: Structuring Organizations to Elicit Guilt as an Affective Reaction to Failure. *Organization Science*, 24(4), 1157–1173.
- Buckert, M., Schwieren, C., Kudielka, B. M., & Fiebach, C. J. (2017). How stressful are economic competitions in the lab? an investigation with physiological measures. *Journal of Economic Psychology*, 62, 231–245.
- Buser, T. (2016). The impact of losing in a competition on the willingness to seek further challenges. *Management Science*, 62(12), 3439–3449.
- Buser, T., & Dreber, A. (2016). The flipside of comparative payment schemes. *Management Science*, 62(9), 2626–2638.
- Cahlíková, J., & Cingl, L. (2017). Risk preferences under acute stress. *Experimental Economics*, 20(1), 209–236.
- Carver, C. S., & Scheier, M. F. (1990). Origins and Functions of Positive and Negative Affect: A Control-Process View. *Psychological Review*, 97(1), 19–35.

- Cassar, L., & Klein, A. H. (2019). A Matter of Perspective: How Failure Shapes Distributive Preferences. *Management Science*, 65(11), 5050–5064.
- Certo, S. T., Busenbark, J. R., Woo, H. S., & Semadeni, M. (2016). Sample selection bias and Heckman models in strategic management research. *Strategic Management Journal*, 37(13), 2639–2657.
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 385–396.
- Cohen, S. (1980). Aftereffects of stress on human performance and social behavior: A review of research and theory. *Psychological Bulletin*, 88(1), 82–108.
- Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journals of Economic Perspectives*, 27(3), 79–102.
- Crocker, J., & Park, L. E. (2004). The costly pursuit of self-esteem. *Psychological Bulletin*, 130(3), 392–414.
- Crocker, J., & Wolfe, C. T. (2001). Contingencies of self-worth. *Psychological Review*, 108(3), 593–623.
- Daviu, N., Bruchas, M.R., Moghaddam, B., Sandi, C., & Beyeler, A. (2019). Neurobiological links between stress and anxiety. *Neurobiology of Stress*, 11.
- Delaney, L., Fink, G., & Harmon, C. (2014). Effects of stress on economic decision-making: Evidence from laboratory experiments. *IZA Discussion Paper*, 8060.
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute Stressors and Cortisol Responses: A Theoretical Integration and Synthesis of Laboratory Research. *Psychological Bulletin*, 130(3), 355–391.
- Giannakakis, G., Pediaditis, M., Manousos, D., Kazantzaki, E., Chiarugi, F., Simos, P. G., Marias, K., & Tsiknakis, M. (2017). Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and Control*, 31, 89–101.
- Gill, D., & Prowse, V. (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics*, 5(2), 351–376.
- Goette, L., Bendahan, S., Thoresen, J., Hollis, F., & Sandi, C. (2015). Stress pulls us apart: Anxiety leads to differences in competitive confidence under stress. *Psychoneuroendocrinology*, 54, 115–123.
- Habhab, S., Sheldon, J. P., & Loeb, R. C. (2009). The relationship between stress, dietary restraint, and food preferences in women. *Appetite*, 52(2), 437–444.
- Haushofer, J., Jang, C., Lynham, J., & Abraham, J. (2018). Stress and temporal discounting: Do domains matter? *Mimeo*.
- Haushofer, J., Chemin, M., Jang, C., & Abraham, J. (2020). Economic and psychological effects of health insurance and cash transfers: Evidence from a randomized experiment in Kenya. *Journal of Development Economics*, 144, 102416.
- Haushofer, J., Jain, P., Musau, A., & Ndeti, D. (2021). Stress may increase choice of sooner outcomes, but not temporal discounting. *Journal of Economic Behavior & Organization*, 183, 377–396.
- Heatherton, T. F., & Polivy, J. (1991). Development and Validation of a Scale for Measuring State Self-Esteem. *Journal of Personality and Social Psychology*, 60(6), 895–910.
- Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153–161.
- International Cognitive Ability Resource Team (2014). The International Cognitive Ability Resource.
- Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The ‘Trier Social Stress Test’ - A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*, 28, 76–81.
- Kudielka, B. M., Buske-Kirschbaum, A., Hellhammer, D. H., & Kirschbaum, C. (2004). HPA axis responses to laboratory psychosocial stress in healthy elderly adults, younger adults, and children: Impact of age and gender. *Psychoneuroendocrinology*, 29(1), 83–98.
- Leder, J., Häusser, J. A., & Mojzisch, A. (2015). Exploring the underpinnings of impaired strategic decision-making under stress. *Journal of Economic Psychology*, 49, 133–140.
- List, J. A., Shaikh, A. M., & Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, 22(4), 773–793.
- Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, 31(3), 301–306.
- Newman, E., O'Connor, D. B., & Conner, M. (2007). Daily hassles and eating behaviour: The role of cortisol reactivity status. *Psychoneuroendocrinology*, 32(2), 125–132.
- Oliver, G., Wardle, J., & Gibson, E. L. (2000). Stress and food choice: A laboratory study. *Psychosomatic Medicine*, 62(6), 853–865.
- Petzold, A., Plessow, F., Goshcke, T., & Kirschbaum, C. (2010). Stress reduces use of negative feedback in a feedback-based learning task. *Behavioral Neuroscience*, 124(2), 248.

- Porcelli, A. J., & Delgado, M. R. (2017). Stress and decision making: Effects on valuation, learning, and risk-taking. *Current Opinion in Behavioral Sciences*, 14, 33–39.
- Raio, C. M., Hartley, C. A., O'Donoghue, T. A., Li, J., & Phelps, E. A. (2017). Stress attenuates the flexible updating of aversive value. *Proceedings of the National Academy of Sciences*, 114(42), 11241–11246.
- Rutters, F., Nieuwenhuizen, A. G., Lemmens, S. G., Born, J. M., & Westerterp-Plantenga, M. S. (2009). Acute stress-related changes in eating in the absence of hunger. *Obesity*, 17(1), 72–77.
- Shields, G. S., Szalma, M. A., & Yonelinas, A. P. (2016). The effects of acute stress on core executive functions: A meta-analysis and comparison with cortisol. *Neuroscience & Biobehavioral Reviews*, 68, 651–668.
- Šidák, Z. (1967). Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association*, 62(318), 626–633.
- Starcke, K., Polzer, C., Wolf, O. T., & Brand, M. (2011). Does stress alter everyday moral decision-making? *Psychoneuroendocrinology*, 36(2), 210–219.
- Starcke, K., Wiesen, C., Trotzke, P., & Brand, M. (2016). Effects of acute laboratory stress on executive functions. *Frontiers in Psychology*, 7, 461.
- Starcke, K., Wolf, O. T., Markowitsch, H. J., & Brand, M. (2008). Anticipatory stress influences decision making under explicit risk conditions. *Behavioral Neuroscience*, 122(6), 1352.
- Stevens, A. H. (1999). Climbing out of poverty, falling back in: Measuring the persistence of poverty over multiple spells. *The Journal of Human Resources*, 34(3), 557–588.
- Sunde, U., & Dohmen, T. (2016). Aging and preferences. *The Journal of the Economics of Ageing*, 7, 64–68.
- Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A. R., & Updegraff, J. A. (2000). Biobehavioral responses to stress in females: tend-and-befriend, not fight-or-flight. *Psychological Review*, 107(3), 411–429.
- Thaler, R. H., & Johnson, E. J. (1990). Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice. *Management Science*, 36(6), 643–660.
- Vitt, N., James, J., Belot, M., & Vecchi, M. (2021). Daily stressors and food choices: A lab experiment with low-SES mothers. *European Economic Review*, 136, 103754.
- Von Dawans, B., Fischbacher, U., Kirschbaum, C., Fehr, E., & Heinrichs, M. (2012). The Social Dimension of Stress Reactivity: Acute Stress Increases Prosocial Behavior in Humans. *Psychological Science*, 23(6), 651–660.
- Von Dawans, B., Kirschbaum, C., & Heinrichs, M. (2011). The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format. *Psychoneuroendocrinology*, 36(4), 514–522.
- Wölbert, E., & Riedl, A. (2013). Measuring time and risk preferences: Reliability, stability, domain specificity. *CESifo Working Paper*, 4339.
- Wooldridge, J.M. (2010). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press, 2nd ed.
- Zellner, D. A., Loaiza, S., Gonzalez, Z., Pita, J., Morales, J., Pecora, D., & Wolf, A. (2006). Food selection changes under stress. *Physiology & Behavior*, 87(4), 789–793.
- Zellner, D. A., Saito, S., & Gonzalez, J. (2007). The effect of stress on men's food selection. *Appetite*, 49(3), 696–699.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.