Oxford Economic Papers, 2025, 1–16
https://doi.org/10.1093/oep/gpaf008
Article

OXFORD

Performance-related pay and sorting into stress

Nicole Andelic [1], Julia Allan [2], Keith A. Bender [1,*], Daniel Powell [3], Ioannis Theodossiou [4]

[1]Department of Economics, Centre for Labour Market Research, and Scottish Experimental Economics Laboratory, University of Aberdeen, Edward Wright Building, Dunbar Street, Aberdeen, AB24 3QY, United Kingdom
[2]Department of Psychology, University of Stirling, Stirling, FK9 4LA, United Kingdom
[3]Health Psychology, Institute of Applied Health Sciences, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, United Kingdom
[4]Department of Economics, Centre for Labour Market Research, University of Aberdeen, Edward Wright Building, Dunbar Street, Aberdeen AB24 3QY, United Kingdom

*Corresponding author. Department of Economics, Centre for Labour Market Research, and Scottish Experimental Economics Laboratory, University of Aberdeen, Edward Wright Building, Dunbar Street, Aberdeen AB24 3QY, United Kingdom. E-mail: kabender@abdn.ac.uk

Abstract

To date, the literature on the performance-related pay (PRP) and health relationship has focused on self-reported data. This article uses an experimental method to examine the effect of PRP on stress measured by salivary cortisol for those who self-selected into a PRP contract, focusing on whether participants who perceive themselves with higher ability self-select into PRP and exhibit different stress changes compared to those not in a PRP scheme. Results show that self-selected PRP participants demonstrate significantly higher cortisol levels than participants in the non-PRP group. This study suggests that, regardless of sorting, PRP leads to higher physiological stress.

Keywords: performance-related pay; stress; experiment; cortisol
JEL classifications: J33, I0, C91

1. Introduction

To date, much of the economics literature on performance-related pay (PRP) has focused on the advantages of individuals sorting themselves into various forms of PRP to increase worker productivity. However, in the book, *Wealth of Nations*, the Scottish economist Adam Smith noted that '*Workmen … when they are liberally paid by the piece, are very apt to overwork themselves and to ruin their health and constitution in a few years*' (Smith 1937: 83). In recent years, a number of studies have found empirical evidence indicating PRP leads to poor health. PRP workers report more injuries (Bender, Green, and Heywood 2012), poorer cardiovascular and digestive health (Bender and Theodossiou 2014), and worse mental health (Dahl and Pierce 2020). Although much of this research has relied on self-reported data (or subjective data as in Dahl and Pierce 2020) which can be subject to biases. These biases arise from two sources. First, self-reported and subjective responses can be subject to non-random misreporting, cognitive dissonance and similar biases. Second, the effect of PRP on health can be influenced by self-selection effects as it can be argued that workers who are attracted to PRP have more preferences for risk compared to their non-PRP counterparts (Grund and Sliwka, 2010; Cornelissen, Heywood, and Jirjahn 2011; Bandiera et al. 2015). As Cobb-Clark, Dahmann, and Kettlewell (2022), Anderson and Mellor (2008), and Dohmen et al. (2011) argue, this correlates with risky behaviours which in turn are correlated with poorer mental and physical health. Thus, these workers with high preference for risk are more likely to be found in PRP contracts and to exhibit poorer health. Another line of argument in the literature is that workers who sort themselves into PRP may have higher cognitive ability (Curme and Stefanec 2007), which is also correlated with better health (Cutler and Lleras-Muney 2010). Hence, workers who self-select into PRP contracts could appear to enjoy better health compared to their non-PRP counterparts. Since in survey data random assignment is very rare, self-selection biases contaminate estimates, regardless of whether it is improving or worsening health.

To confront the above inadequacies of the relevant literature, Allan, Bender, and Theodossiou (2020) and Allan et al. (2021) use experimental methodology in an economics laboratory setting where participants are randomly allocated to PRP and non-PRP regimes. By randomly allocating participants to either a PRP or a fixed payment regime, the authors circumvent the issue of self-selection. The studies find that even a brief 10-min work task under PRP leads to higher levels of the stress-related hormone, cortisol, in saliva when participants are paid by performance rather than a fixed payment.

Although Allan, Bender, and Theodossiou (2020) and Allan et al. (2021) find evidence of a strong causal relationship between PRP and physiological stress, their use of randomization also removes participant decision-making from the process. The question then arises whether in an experimental laboratory set up sorting has the expected biases as in the survey data. Do the data generated in the experimental laboratory replicate the repercussions of sorting as in the real labour market when the workers can choose the contract type that suits them best? Is it the case that the same factors that lead to individuals sorting into PRP contracts may also cause them to be more resilient to stress? Does sorting into PRP mitigate some of the acute stress caused by PRP in an experimental setting? Although there are a few experimental studies investigating PRP and stress, to date there are none which include both sorting and robust measures of stress. The current study aims to address this gap by investigating the impact of sorting into PRP on physiological stress, to assess whether self-selection at all mitigates the stress that has been shown to be generated by PRP in earlier studies.

2. Brief literature review

One of the main arguments in support of PRP is the belief that PRP allows for the alignment between the interests of the firm and the workers. There is some evidence for PRP being beneficial for workers: using panel survey data from the British Household Panel Survey, Green and Heywood (2008) find an association between PRP and overall job satisfaction including satisfaction with pay, hours and job security. On the other hand, using the same data McCausland, Pouliakas, and Theodossiou (2005) demonstrate that the association is moderated by income, suggesting that low-income workers on PRP contracts do not find the job as satisfying as those who are highly paid. The benefit of PRP for firms is more robust; there is a large literature showing evidence of increased productivity when using performance pay contracts. The reason for the increase in productivity is two-fold. Firstly, offering performance pay may incentivize employees to work harder and therefore produce more (Lazear 2000). This association between productivity and PRP has been found both when observing firms where there has been an exogenous shift from performance pay contracts to fixed salaries (Lazear 2000; Nagin et al. 2002; Freeman and

Kleiner 2005) and in laboratory settings (Dohmen and Falk 2011; Cadsby, Song, and Tapon 2016). In summary, PRP is consistently found to increase effort and therefore output. Secondly, the literature argues that PRP contracts attract more productive workers through a sorting mechanism (Lazear 2000). Workers who are highly productive seek PRP jobs in order to maximize their earnings whereas workers who are less productive seek fixed salary jobs (Lazear 2000). Indeed, Lazear (2000) estimates that 56 per cent of the productivity increase associated with PRP is due to this sorting effect. However, firms switching payment schemes can rarely be considered a completely exogenous event. Furthermore, the firm in Lazear's study (2000) implemented a minimum wage for their PRP workers, meaning that workers were guaranteed to be paid either the same as before or more under the new piece rate scheme. Consequently, the 'risk' of not earning enough in PRP contract was removed in this instance. Finally, in real world data it is often difficult to find information on individual performance under both payment schemes. To control for these circumstances then it may be necessary to use experimental methods instead.

One such example is a laboratory study presented by Eriksson and Villeval (2008), in which they carry out an experiment with two stages; (1) a fixed wage condition and (2) a choice between a fixed and variable wage condition. The authors find that low-skill participants are more likely to remain in the fixed wage condition, whereas high-skill workers switch to a variable payment scheme when given the opportunity where they subsequently put in more effort than in the fixed wage condition. However, the same study finds that if firms offer higher fixed wages, output increases in the fixed payment condition and high-skill participants are less likely to switch conditions. These results suggest that productivity is not the only factor that determines whether high-performing individuals select variable or fixed pay.

Another factor which may impact on payment contract preference is the extent to which workers are risk-averse. In an early survey study, Cable and Judge (1994) show that workers who are highly risk-averse are more likely to prefer fixed pay over variable pay. In line with this preference, Shearer and Bellemare (2006) find in a field study that workers in a tree-planting firm who are paid by piece rate are significantly more risk-tolerant than individuals from the general population, when taking part in low- and high-stake lotteries. These findings have been replicated in various lab experiments, albeit the relationship may be less straightforward than suggested by earlier research. For example, Cadsby, Song, and Tapon (2016) suggest that the relationship between risk-aversion and PRP-selection is moderated by capability. Thus, individuals who are risk-averse yet estimate high performance are still more likely to choose PRP than those who estimate low performance even if PRP generates an uncertain payout. Furthermore, Cadsby, Song, and Tapon (2016) argue that this may interact with stress; risk-averse individuals may feel more stress during PRP payment conditions, which, in turn, impairs their performance. These studies confirm that PRP workers indeed differ in risk aversion from fixed salary workers. Importantly though, none of these studies include a measure of stress.

2.1 Pathways between PRP and health

Regardless of how workers sort into PRP, variable pay has repeatedly been linked with poorer health in studies based on survey data. There are several different pathways between PRP and health. Firstly, PRP may lead to higher rates of accidents and injuries. Through incentivizing higher production, workers paid by PRP are encouraged to take more risks. For example, Freeman and Kleiner (2005) find that workers' compensation insurance premiums decreased after the firm switched from piece rate pay to fixed salaries, suggesting that injuries decreased after the change in salaries. Using European survey data, Bender, Green, and Heywood (2012) also show a strong relationship between piece rates and injuries at work. Secondly, if workers are incentivized to produce more, they may also work longer hours which reduces the time available for healthy behaviours such as sleep and exercise. Bender and Theodossiou (2014) use a large-scale survey that shows that PRP employees work on average 1.5 h more per week than those in fixed contracts. They also find that the longer workers are in PRP contracts, the more likely they are to suffer from heart, stomach and other health problems. It is also possible that PRP leads to behaviours which are detrimental for health. Again, Bender and Theodossiou (2014) show that PRP workers are likely to drink more alcohol, and Artz, Green, and Heywood (2021) report that alcohol/drug use is more common in PRP workers than workers with fixed pay. Finally, the fluctuations in the income stream associated with PRP may be inherently more stressful, causing 'allostatic load' to increase over time (McEwen 1998). This persistent stress to the body can compromise the immune system and increases the likelihood of other health problems.

These studies establish a robust association between PRP and health, with multiple potential pathways. As stress is one potential mechanism, we argue that there is a need for further research on the causal relationship between PRP and stress. Experimental studies are useful in this instance. First, Allan et al. (2021) use an experimental design to indicate the causal effect of PRP on stress. Participants were invited to an experimental session where they were randomly allocated to either PRP or minimum performance fixed payment conditions. A week after the first experiment, participants were invited back to take part in a follow-up session, during which their payment contract switched from their original session. This crossover design allowed for participants to serve as their own controls. In addition to self-reported stress measures, the authors also collected samples of salivary cortisol, a hormone highly sensitive to acute stress (Dickerson and Kemeny 2004). Overall, the results show that when participants take part in the randomly allocated PRP condition they do not recover from the stressful work task as quickly as those who take part in the minimum performance condition. The randomized crossover design allows the authors to control for individual effects when measuring the effect of PRP on stress. However, although the design allows the study to establish a much-needed causal relationship between PRP and stress, it does not allow us to draw any conclusions on how self-selection may mitigate the potential stress felt in PRP contracts.

Dohmen and Falk (2011), on the other hand, uses experimental methods to that allow participants to self-select into PRP or fixed payment contracts for carrying out a math task to study the effect of PRP on productivity. Although it was not the main purpose of the study, they also find that individuals in the PRP condition self-report higher levels of stress than those in the fixed payment condition. However, the study suffers from some limitations. Firstly, as stress is a secondary objective of their study, Dohmen and Falk did not measure stress in participants before they had taken part in the task. It is therefore not possible to rule out that those who self-selected into PRP were already more stressed than the fixed payment group prior to taking part in the work task. Furthermore, the nature of self-reported data means that it is possible that those who self-selected PRP were subject to biases when self-reporting their stress. Finally, Dohmen and Falk paid their PRP subjects a fee per correct answer but did not require their fixed payment subjects to complete any calculations to earn their fixed salary. Even firms that implement a fixed wage usually require a minimum level of output by their workers (Lazear 2000). The difference in stress then may be driven by participants in the fixed payment condition knowing that there was no risk of not receiving a payment.

2.2 Sorting mechanism

Before turning to the differences in experimental design between this paper and Dohmen and Falk (2011), it is important to note that the sorting mechanism into PRP is key to both papers. Dohmen and Falk (2011) identify a number of factors that lead to sorting into PRP (particularly piece rates which is the focus of the experiment explained below), namely attitude towards risk and expected productivity. It is clear to see that these will affect the potential performance in the task as found by Dohmen and Falk (2011). However, it is likely to also have an effect on stress. Those who prefer more risky situations in particular will either have innate ways of or develop coping strategies to mitigate the stress of those risky situations. Thus, one would expect only a muted or maybe no stress response at all in a risky PRP contract. Indeed, as discussed below, the participants recruited for this experiment show similar sorting behaviours over risk preferences and productivity into PRP, and thus, the question for this study is whether these preferences insulate PRP participants from the stress of these jobs when they are allowed to select into a PRP contract.

2.3 The current study

From the summary of the two key experimental papers above, we can conclude that there is a gap in the literature for an experimental study examining the role of self-selection in PRP and stress (which is a key feature of the labour market) but that also includes more robust measures of stress, before and after the work task. The current study then is a combination of the two studies described in the previous section. In line with Dohmen and Falk (2011), participants are invited to an experimental study in which they self-select into their payment contract like workers do in the labour market. This is contrary to Allan et al. (2021) which forces participants into a contract that they may not want to be in, which in itself may increase their stress. However, we differ from Dohmen and Falk (2011) in two important respects. First and most importantly, salivary cortisol is measured before and after the work task, as is done in the study by Allan et al. (2021). Second, the non-PRP condition in Dohmen and Falk (2011) had no requirement to work at all—there was no minimum performance needed to earn the flat fee, which does not reflect behaviours in labour markets and may cause an inaccurate comparison between PRP and a completely stressless non-PRP/fixed payment condition. The aim of the experiment, therefore, is to examine the effect of sorting into performance related or fixed payment conditions on both self-reported stress and salivary cortisol.

3. Methods

One-hundred and fifty-five participants were recruited through an internal recruitment platform for experimental economics, consisting of undergraduate and postgraduate students potentially from any academic discipline. Prior to attending, participants were advised that they would have to take part in two sessions on the same day: a short practice session in the morning and then an incentivized session that same afternoon. The inclusion of a practice session was motivated by an unpublished earlier study which found that not including a practice session led to high levels of overconfidence in likely performance and ability. Consequently, a morning session was included to ensure that participants were familiar with their ability, as in the labour market. Participants were paid £10 as payment for participation with an opportunity to earn more during the experiment. Ethical approval was granted by the University of [redacted] School of [redacted] (CERB/2019/12/1831). The experiment took place over 22 sessions between March and May 2022. To control for the diurnal pattern of cortisol (Czeisler and Klerman 1999), every afternoon session took place at 2 pm on weekdays. As cortisol is affected by a range of external factors, participants were instructed to refrain from eating, alcohol or nicotine use, brushing their teeth or engaging in strenuous physical activity in the two hours immediately before the experimental session. Upon arrival participants were randomly allocated to one of 20 PCs running the experiment via the experimental software z-Tree (Fischbacher 2007).

Before describing the experimental session, it is worth describing the main part of the experiment—namely the nature of the contracts from which participants can choose. The PRP contract is for £0.20 per correct answer to a set of math problems while the non-PRP contract is a minimum performance contract generating a flat payment of £5 if there are ten correct answers. There are a couple reasons why this particular set of characteristics are chosen for this experiment. First, as discussed in the introduction, risk is an important element here and so having an element of risk of not hitting the non-PRP payment for those who might choose the PRP contract is a central concern. If the piece rate were such that it would generate £5 for ten correct answers as in the non-PRP contract, this would mean a very large piece rate (£0.50). Second, the choice of piece rate should also generate a relatively equal split between the choice of the two contracts so that there is variation in the sorting. For a very low piece rate, most would choose the safe non-PRP contract while if the piece rate is very high, then most would choose the PRP contract. Pre-testing suggested that having a piece rate which targeted twenty to twenty-five correct answers for the PRP to equate to the money needed to generate the same as in the non-PRP contract was adequate to have sufficient variation. One would be concerned if participants were effort avoiders there would be very few who would choose the PRP contract, but as mentioned below, 55 per cent of participants chose the PRP contract—a ratio similar to the proportion that choose the piece rate contract of Dohmen and Falk (2011). Of course, the choice of these PRP characteristics may not be representative of some actual PRP schemes offered by firms or even the more theoretical discussions of sorting and PRP found in papers like Lazear (2000), but it allows for comparisons of the stress reactions between the two groups after sorting given the relatively equal split into PRP and non-PRP contracts.

The practice session was always scheduled for 10.30am and typically lasted for approximately 30 min. After being randomly seated, participants carried out a 10 min-work task consisting of math calculations under two hypothetical payment contracts: a PRP contract during which they would be paid £0.20 per correct answer or a non-PRP contract of £5 for having solved at least ten problems correctly during the work task. The order in which the contracts were presented was randomized across participants. After completing both tasks, participants were informed of their performance and their hypothetical earnings under each contract. This was designed to allow them to make an informed choice about which payment condition they would select into in the later main session. They were then asked to complete a questionnaire (Survey 1) where they rated their willingness to take risks from 1 to 7 (‘not at all’, ‘greatly’), their perceived arithmetic skill and resilience to stress from 1 to 5 (‘very well’ to ‘very poor’) as well as demographic questions. Before leaving, they were reminded to return in the afternoon and to avoid activities that could influence cortisol levels as salivary samples were part of the afternoon session.

As in Allan et al. (2021) and shown in Fig. 1, each afternoon session consisted of four 10-min phases to allow sufficient time for changes in cortisol to occur (Kirschbaum, Pirke, and Hellhammer 1993). The first phase, Phase 1, consisted of 10 min of relaxation, during which subjects were provided with colouring-in sheets designed to encourage mindfulness

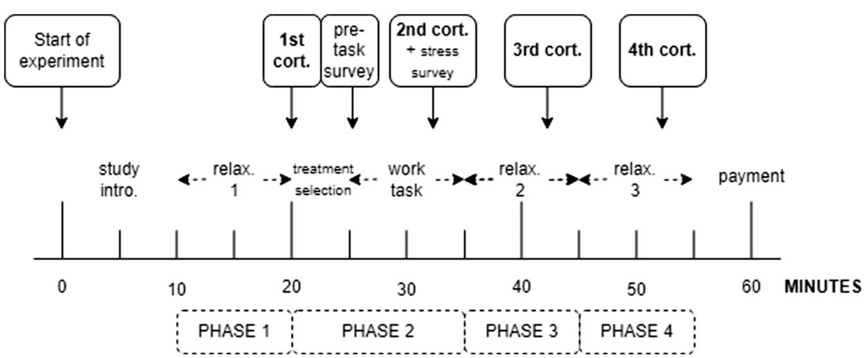

Figure 1. Experimental procedure of afternoon session.

and coloured pencils. In the second phase, they were then given the following brief description:

> 'Option A: You will be paid £5 for having solved at least 10 problems. You will not be paid anything if you answer less than 10 questions correctly, and you will not be paid any extra for more than 10 correct answers.'
>
> 'Option B: You will be paid £0.20 for each correctly solved mathematical problem. You will not be paid anything for an incorrect answer.'

Participants were then prompted to choose one of the conditions and then to provide the first cortisol sample. To provide salivary cortisol, participants chewed on a synthetic swab (Cortisol Salivettes® (Sarstedt, Nümbrecht, Germany)) for 60 s and then placed it in a labelled test tube. They then completed a second survey (Survey 2) which contained two subjective stress items ('How stressed are you right now?' and 'How exhausted are you right now?') rated on a scale from 1 to 4 ('not at all' to 'much more than usual') as well as six items assessing their choice of payment contract:

'Before the task I feel … ' (1) 'that my choice is going to be more enjoyable than the alternative', (2) 'that my choice is going to be more challenging than the alternative', (3) 'that my choice is going to be more stressful than the alternative', (4) 'that my choice is going to be more financially lucrative than the alternative', (5) 'that I am satisfied with my choice', and (6) 'that I regret my choice'. All were rated from 1 to 5 ('highly agree' to 'highly disagree'). As an additional measure of risk preference, participants were then asked to make the following choice: 'Imagine you were entering one of two lotteries. Which lottery would you choose?'

> Option 1 (Lower Risk Lower Reward): 'You have a 50% chance of winning £10'
>
> Option 2 (Higher Risk Higher Reward): 'You have a 25% chance of winning £20'

Next, participants were presented with a list of activities and medications that produce known physiological changes in cortisol and asked to indicate if any had been used, so that these could be controlled for in the analyses. Finally, participants completed the main work task.

As in Dohmen and Falk (2011) and Allan et al. (2021), the work task asked participants to complete randomly generated mathematical questions by hand over 10 min. There were a maximum of fifty questions generated, and although the values were randomized across sessions and participants, the ratio of questions (twenty multiplications, ten divisions, and ten additions) remained the same for all participants to ensure that the difficulty level of the task was approximately consistent for the full sample. The number of correct answers and time left to complete the work task were visible in the top corner of the screen. When participants in the minimum performance condition had achieved the required ten correct answers, a banner at the top of the screen informed them that they had earned an additional £5, but that they were free to continue solving questions for the remainder of the time if they wished. After the work task was done, participants received their results and a summary of their earnings. Then they completed a second cortisol sample and a third survey (Survey 3), consisting of the following four items rated on a four-point scale (ranging from 'not at all' to 'very' or 'great'):

> 'After the task, how stressed do you feel?'
>
> 'After the task, how exhausted do you feel?'

'How much effort did exert solving the mathematical problems in the previous 10 minutes?'

'Did you feel under strain when solving the mathematical problems in the previous 10 minutes?'

The survey also included the arithmetic ability item and the six items about the payment contract choice from Survey 2. Finally, participants were asked if they would choose the same contract if they were to complete the task again and to briefly explain the motivation for their decision.

Phase 3 and 4 both consisted of 10 min of relaxation followed by a third and fourth cortisol sample, respectively. Participants were then paid in a separate room before leaving the experiment. All cortisol samples were frozen at −20°C on the day of collection. After ten weeks when the study had been completed all samples were shipped to Biochemisches Laboratory, University of Trier, to analyse cortisol levels.

4. Results

4.1 Descriptive statistics

One-hundred and fifty-five participants took part in the second experiment. As is common in the cortisol literature (Nicolson 2008), participants with cortisol levels more than four standard deviations away from the mean were excluded from the analysis, leaving 147 participants. As cortisol levels are sensitive to daily activities such as recent food intake, caffeine, nicotine, physical activity, diurnal rhythm and some medications, each participant was provided with a list in which they ticked any activities or medications which may affect cortisol levels. Participants who disclosed such confounders were then coded with a dummy variable capturing confounding activities or confounding medications depending on their disclosure. The majority of the sample are female, between 21 and 23 years old, postgraduate students and studying in the School of Arts & Social Sciences (see Table 1). Interestingly there are only minor differences in the characteristics across the PRP and non-PRP samples.

4.2 Predictors of sorting

The first step of the analysis was to examine the choice between PRP and non-PRP. There was a preference for PRP over a fixed payment (55.10 vs 44.90 per cent), in line with the

Table 1. Sample characteristics.

	Full sample $n = 147$	Non-PRP $n = 66$	PRP $n = 81$
Male	56 (38%)	25 (38%)	31 (38%)
Age 18–20	28 (19%)	13 (20%)	15 (19%)
Age 21–23	59 (40%)	28 (42%)	31 (38%)
Age 24–26	22 (15%)	9 (14%)	13 (16%)
Age 27–29	17 (5%)	7 (11%)	10 (12%)
Age 30+	21 (14%)	9 (14%)	12 (15%)
1st year	23 (16%)	9 (14%)	14 (17%)
2nd year	23 (16%)	11 (17%)	12 (15%)
3rd year	36 (24%)	15 (23%)	21 (26%)
4th year	25 (17%)	14 (21%)	11 (14%)
Other year	40 (27%)	17 (26%)	23 (28%)
Arts & Social Sciences	44 (30%)	22 (33%)	22 (27%)
Business	41 (28%)	14 (21%)	27 (33%)
Life Sciences	34 (23%)	18 (27%)	16 (20%)
Physical Sciences	28 (19%)	12 (18%)	16 (20%)

Source: Authors' calculations.

findings documented by Dohmen and Falk (2011) where 61 per cent of their sample chose the PRP contract. The decision about payment contract was well-paid for the majority of the participants and more than three quarters reported that their decision would remain the same if given the opportunity to do the experiment again (76.87 per cent). To investigate who chose the PRP condition, a binomial logistic regression predicting contract choice is estimated finding that none of the socio-demographic (gender, age, discipline, or year of study) were significant predictors. Instead, achieving a higher number of correct answers during the practice task (OR = 1.13) is the only statistically significant predictor of choosing PRP. These findings suggest that performance on the practice task guided participants towards choosing the treatment which offered the possibility of earning more. To a certain extent then, the findings support Lazear's (2000) argument that if workers think that they can earn more in PRP, they will choose variable pay over the fixed salary.

Unsurprisingly, participants who selected into non-PRP perceive themselves as poorer at maths than those selecting into PRP (PRP: 2.23, non-PRP: 2.72, $t = 2.98$, df = 144, $P = 0.003$) echoing findings in Dohmen and Falk (2011), but that there is no difference in self-reported resilience between the groups (2.62 vs 2.73). In contrast to previous studies (Cable and Judge, 1994; Shearer and Bellemare 2006) there is no difference in self-reported risk willingness (4.79 vs 4.67), although PRP participants are more likely to choose the risky lottery option (32.10 per cent) than the non-PRP group (18.18 per cent). However, this association is only significant at a 10% level ($X^2 = 2.98$, df = 1, $P = 0.084$).

Finally, as Cadsby, Song, and Tapon (2016) find that capability moderated the relationship between choosing PRP and risk tolerance, the same model is tested here using the lottery task as our measure of risk. As before, performance in the practice session independently predicts the probability of choosing PRP (0.19), but there is no independent effect of the lottery task as a risk measure nor an interaction effect between the two. The results remain qualitatively the same after controlling for socio-demographic variables.

4.3 The effects of PRP on stress

4.3.1 Stress measured via self-report

There is no difference in stress (PRP: 2.14, non-PRP: 2.12) between the PRP and fixed payment groups measured prior to the task. After the task PRP participants report a significantly higher increase in stress (0.33 vs −0.03, $t = -2.32$, df = 145, $P = 0.022$), higher self-reported strain (3.01 vs 2.36, $t = -4.41$, df = 145, $P < 0.001$) and effort (3.46 vs 3.12, $t = -3.07$, df = 145, $P = 0.003$) in comparison to the fixed payment group. These differences remain even after controlling for socio-demographic characteristics. Surprisingly, exhaustion prior to the work task is slightly higher among non-PRP participants (2.12 vs 2.41, $t = 1.94$, df = 145, $P = 0.055$), but this difference does not remain after the task (2.28 vs 2.33).

4.3.2 Stress measured via cortisol

Higher levels of cortisol indicate higher levels of stress. As outlined in the methods section, to allow ten minutes for cortisol changes to be reflected in the saliva (Kirschbaum, Pirke, and Hellhammer 1993), participants provided four measures of cortisol throughout the task: a baseline, an immediate post-task, 10 min post-task, and 20 min post-task sample. Due to natural individual variation in cortisol levels, the focus of this study is on changes in cortisol over time between the baseline samples and the post-task samples rather than on levels of cortisol per se. This change in cortisol can be measured in multiple ways, and so the measures reported here are 'Area under the Curve with respect to increase' (AUCi), 'peak change' in cortisol (the difference between the peak value out of the three post-task samples and the baseline sample), and 'overall change' in cortisol (the difference between the final +20 minutes sample and the baseline sample). A positive outcome value indicates higher cortisol post-task and a negative value indicates lower cortisol post-task. Finally, an

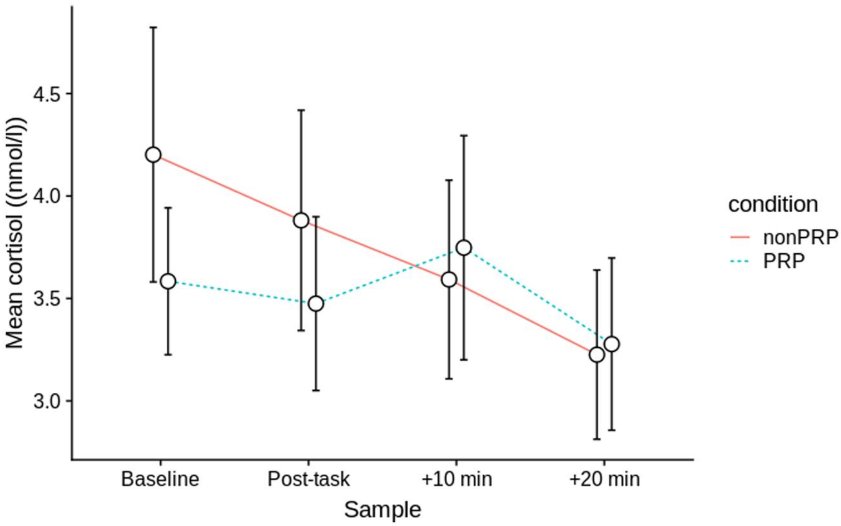

Figure 2. Raw cortisol levels with 95% CI.

overall cortisol output measure, ‘Area Under the Curve with respect to ground’ (AUCg), was also calculated (see Pruessner et al. 2003). Throughout all analysis, peak change was log-transformed to address positively skewed data. The remaining variables (AUCi, overall change and AUCg) were not transformed.

Although the focus is on the change in cortisol over the duration of the experiment, for completeness, raw cortisol samples are also examined. Figure 2 shows that, although non-PRP participants demonstrate slightly higher baseline cortisol levels than the PRP group (4.20 vs 3.58, $t = 1.76$, df = 145, $P = 0.080$),[1] there is no significant difference in Sample 2 (3.88 vs 3.47), Sample 3 (3.59 vs 3.75) or Sample 4 (3.23 vs 3.28).

Consistent with the self-reported stress results, one-sided t-tests show significantly higher levels of AUCi (−1.53 vs −15.78, $t = -1.87$, df = 145, $P = 0.031$), (log of) peak change (2.33 vs 2.28, $t = -2.02$, df = 145, $P = 0.023$)2, and overall change (−0.31 vs −0.98, $t = -2.16$, df = 145, $P = 0.016$) in the PRP groups. There is no significant difference in AUCg between the groups (141.82 vs 152.28).

As a robustness check, several OLS regressions are estimated for each dependent variable to allow adjustment for the presence of known cortisol confounders and demographic situational factors. The first regression includes only known cortisol confounders as covariates (e.g. recent activities or use of medications) whereas the second regression also controls for sociodemographic and contextual variables (sex, age, year, and area of study as well as the first condition experienced in the practice session and whether participants had taken part in lab experiments previously).

[1] Baseline levels for the non-PRP group are similar to those found for the overall sample in Allan et al. (2021). In contrast, the baseline level for PRP here is slightly lower. Although this difference is interesting, it is not clear if this is due to confidence/absence of anticipation effect among PRP participants or if this is due to natural individual cortisol variation which can be quite substantial (Hruschka, Kohrt, and Worthman 2005). Notably, there is no difference in baseline self-reported stress between the two groups. That there is a marginally significant difference in baseline cortisol might lead to worries about selection into contract type due to levels of cortisol. However, medical research summarized by Starcke and Brand (2016) suggests that there is no consistent correlation between acute measures of stress such as cortisol and risky behaviours. Using the data from the paper’s experiments, there is also no statistical relationship in baseline cortisol and choice of payment contract nor is there a statistical relationship between baseline subjective stress and choice of payment contract in logit regressions conditional on demographic characteristics and inclusion of the performance in the practice session. This is consistent with the medical research reported in Starcke and Brand (2016).

As can be seen in Table 2, PRP remains a significant predictor of higher cortisol change as measured by AUCi (coefficient of 13.62) and (log of) peak change (0.04) at a 10 per cent level and of overall change (0.65) at a 5 per cent level. After including the socio-demographic and context covariates the coefficients remain relatively similar in size but the significance levels are reduced in regressions estimating AUCi (12.06) and overall change (0.57) (to not significant and to a 10% level, respectively). Peak change (0.04) remains unchanged.[2] Of the covariates, being a Business School student (AUCi: 20.42, peak change: 0.05) or being allocated to PRP first in the morning practice sessions (13.66 and 0.05) are significant predictors of higher AUCi and peak change. Although PRP is a statistically insignificant predictor, being male (35.72) and a life science student (41.44) predicts higher AUCg, the only measure not based on changing cortisol.

4.3 PRP and performance

One of the main advantages of PRP is that it leads to higher productivity (Lazear 2000). Indeed, as found in previous experiments by Dohmen and Falk (2011) and Cadsby, Song, and Tapon (2016), PRP subjects achieved a higher number of correct answers (30.51) than non-PRP participants (19.20). In the current study, nineteen PRP participants carried out more than ten but fewer than twenty-five questions correct, suggesting that they would have earned more under the alternate contract. Two non-PRP participants achieved fewer than 10 correct questions and were therefore not eligible for the additional fixed payment of £5 in the non-PRP condition. Despite this, only 14.29 per cent of participants would have earned more had they chosen the alternate contract. This is also a significant increase in performance when comparing this study with the previous experiments using randomized allocation (Allan et al. 2021).

There is no significant relationship between performance and self-reported stress when examining the full sample or either of the conditions separately. Similarly, there is no association between performance and cortisol change.

5. Discussion and future work

The aim of the current study is to examine the relationship between self-selected PRP and stress. It investigates whether participants who perceive themselves as having higher ability in the work task and self-select into PRP exhibit lower or at least equal stress levels compared to those under a non PRP regime. In line with observational survey data, results show that participants generally chose a payment contract based on their perceived task ability, and that PRP participants typically performed better during the practice session than the self-selected fixed payment group had. However, despite this sorting effect, PRP participants were still significantly more stressed than the minimum performance group while completing the main work task.

A common criticism of previous literature is that there may be underlying traits, such as high risk tolerance, which lead to both sorting into PRP as well as poorer health. In previous research, Allan, Bender, and Theodossiou (2020) and Allan et al. (2021) randomly allocated participants to PRP and non-PRP regimes, circumventing the issue of self-selection. However, the use of randomization removes participant decision-making from the

[2] It is interesting to note that despite many of the PRP coefficients are statistically significant only at the 10 per cent level, this is for a two-tailed test. A one-tailed test which examines whether PRP increases cortisol or not is statistically significant at the 5 per cent level for all of the cortisol change dependent variables for the parsimonious regressions that just use the confounders as controls and for overall and (log of) peak change for the regressions with more covariates. Overall, the coefficients are quite small which may be expected for a relatively short laboratory experiment with low stakes. However, it should be noted that perhaps over a longer period, habituation to the task might reduce the effect of PRP on stress, though there is no mechanism to guarantee this, particularly if the stakes are high as would be in the labour market. Interestingly, a recent paper by Andelic et al. (2024) suggests that PRP is associated with other, non-cortisol biomarkers of stress in a nationally representative UK dataset.

Table 2. OLS regression coefficients predicting cortisol.

	AUCi		(log of) Peak change		Overall change		AUCg	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Constant	−12.85 (6.14)	−27.34 (16.23)	2.29 (0.02)	2.22 (0.05)	−0.87 (0.25)	−1.20 (0.67)	155.80 (10.42)	120.64 (28.42)
PRP	13.62 (7.67)	11.69 (7.61)	0.04 (0.03)	0.04 (0.03)	0.65 (0.31)	0.55 (0.31)	−11.68 (13.01)	−7.69 (13.32)
Activities	−8.55 (12.53)	−14.66 (12.94)	−0.02 (0.04)	−0.04 (0.04)	−0.29 (0.51)	−0.43 (0.53)	−1.77 (21.28)	−9.63 (22.65)
Medication	−10.94 (10.49)	−7.72 (10.83)	−0.04 (0.04)	−0.03 (0.04)	−0.39 (0.43)	−0.35 (0.44)	−17.06 (17.81)	−3.55 (18.97)
Male		6.13 (8.77)		0.03 (0.03)		−0.18 (0.36)		35.73 (15.35)
Age 21–23		3.23 (11.90)		0.01 (0.04)		0.39 (0.49)		−20.64 (20.84)
Age 24–26		−15.44 (15.20)		−0.06 (0.05)		−0.46 (0.62)		−32.62 (26.61)
Age 27–29		−6.17 (16.99)		0.001 (0.06)		−0.03 (0.70)		−23.67 (29.75)
Age 30+		13.31 (14.14)		0.04 (0.05)		0.90 (0.58)		−12.09 (24.75)
2nd year		13.51 (13.88)		0.04 (0.05)		0.34 (0.57)		13.59 (24.29)
3rd year		9.51 (13.01)		0.01 (0.04)		0.13 (0.53)		16.46 (22.78)
4th year		0.49 (14.88)		−0.001 (0.05)		−0.22 (0.61)		12.41 (26.09)
Other year		20.44 (14.39)		0.06 (0.05)		0.66 (0.59)		11.25 (25.19)
Business		20.36 (10.24)		0.08 (0.03)		0.60 (0.42)		21.70 (17.92)
Life sciences		−11.60 (10.76)		−0.02 (0.04)		−0.66 (0.44)		41.58 (18.83)
Physical sciences		0.54 (11.96)		0.003 (0.04)		0.12 (0.49)		11.83 (20.94)
PRP first in practice session		13.56 (7.90)		0.05 (0.03)		0.47 (0.32)		17.13 (13.84)
Prior participation		−9.02 (8.93)		−0.01 (0.03)		−0.34 (0.37)		−2.29 (15.63)

Notes: Numbers in parentheses are standard errors. Reference categories are non-PRP, not undertaking activities or using medication, female, age 18–20, 1st year, Arts & Social Sciences, non-PRP first in practice session and no prior participation. Means for each of the dependent variables are: −7.93, 2.31, −0.61, 146.52.
Source: Authors' calculations.

process—a key element of labour markets. The question then arises whether in an experimental laboratory set up there are the same factors at work as in the real labour markets that suggest that individuals self-selecting into PRP contracts are more resilient to stress, hence self-selection into PRP mitigates some of the acute stress caused by PRP in an experimental setting. The current study assesses whether self-selection mitigates the stress that has been shown to be generated by PRP in earlier studies.

In the current study, the experimental methodology allows the focus to be on the immediate change in cortisol (and thereby controlling for individual baselines), suggesting if there is a risk preference, it does not translate into resilience towards stress. Furthermore, a previous counter-balanced experiment measuring PRP and cortisol find that the effect of PRP on cortisol remains even when participants are allocated to both PRP and non-PRP in a randomized order (Allan et al. 2021). There is clearly a sorting effect among PRP workers, but these findings suggest that the sorting effect is primarily through ability and that PRP leads to elevated levels of cortisol even among the self-selected which in turn may impact on overall health over time.

Although the use of experiments allows the examination of the effect of PRP on cortisol in a controlled environment, there are however some common methodological challenges that need to be considered. For example, the current study is limited to only examining the effect of PRP on acute, physiological stress and the financial stakes are relatively low. Thus, the experimental results are most likely biased towards finding no effect, compared to what happens in the labour market. Indeed, PRP workers in the labour market are likely to experience a low-grade but persistent level of stress the longer they work in PRP contracts with real world financial stakes. For example, using data from the British Household Panel Survey, Bender and Theodossiou (2014) find evidence of higher health deterioration in PRP workers than fixed salary workers over time. It is beyond the scope of the current paper to examine how PRP may affect cortisol over longer periods of time, but it would be a fruitful avenue for future research to do so.

Another potential limitation is the difficulty in determining the exact mechanism underlying the link between PRP and stress. For example, to keep the study design as close as possible to the design utilized in Allan et al. (2021), there is a difference in value per question between the non-PRP and PRP groups (£0.50 vs £0.20). It is possible that the stress in the PRP group stems from knowing about that difference. However, in Allan et al. (2021), there is a difference in stress between the two conditions even after the initial session when participants are not aware of the parameters of the alternate contract. The financial value of each solved question is therefore unlikely to be the source of stress here. Another potential mechanism is through the additional effort expanded by the PRP group. However, it is not straightforward to analyse the stress and performance relationship. Stress may lead to lower performance (and vice versa) and statistically controlling for it may cause issues with endogeneity (Cadsby, Song, and Tapon 2016). Furthermore, controlling for performance might not be appropriate as it is likely that at least a portion of the stress experienced by PRP employees in the labour market is indeed due to higher effort.

An extension of the experiment could look in more detail on the relationship between contract choice and stress, when the choice is restricted or not met. For example, a participant may prefer to select into a PRP contract but not be able to get it or vice versa. This would equate to a situation where a firm decides to exogenously change its payment contracts. Not having a preferred contract would likely generate more stress than for those who have their preferred contract, though it would be interesting to see if there are differences in the stress if a PRP preferred worker cannot get a non-PRP contract compared to a worker who prefers a non-PRP contract who is given a PRP contract. Another fruitful avenue of research would be to examine other mechanisms of how PRP might affect stress ruled out by construction here. For example, experiments examining income instability (a common feature of PRP payments), 'real' tasks rather than simply mathematical calculations, differences in the level of piece rates or different performance contracts (e.g. a bonus rather than a piece rate) all would be interesting to examine.

A third area of potential further work would investigate in more detail the 'type' of stress generated by PRP—whether that be the inherent (actual) additional demand, perceived psychological stress, or reactivity involving emotional, cognitive, behavioural, and physiological responses. While only psychological stress has been measured in previous literature (e.g. Dohmen and Falk 2011; Cadsby, Song, and Tapon 2016), both psychological and physiological stress are investigated here. As stress is a multidimensional concept beyond subjective perceptions (Nater 2018), this is an important step forward. The experiments above show that while both psychological stress and physiological stress increase among those under the PRP contract, there is only a weak correlation between the two stress measures. This is not uncommon in the stress literature, with the meta-analysis of lab-based stress tasks by Dickerson and Kemeny (2004) finding robust stress-induced effects on psychological states and cortisol responses but no association between the two measures of stress. There are many potential moderating factors which may explain this, including the presence of diurnal rhythms, the pulsatile nature of cortisol secretion, and underlying chronic stress exposure. Future investigations of stress in PRP would benefit from measuring different facets of the multidimensional stress concept to learn more about any PRP-stress-health pathway.

In summary, the current experiment provides some evidence of PRP leading to small increases in physiological and self-reported stress even if participants are allowed to choose the contract which best matches their abilities. The current study did not find that PRP participants self-reported being more resilient to stress, however, even if this is the case in the labour market, these findings suggest that they are not protected from an increase in cortisol. Whether this effect of PRP would exist in repeated interactions in a real labour market, of course, cannot be determined in a laboratory experiment with low stakes. Taking this result to the field would give a better indication of the longer-run effects of PRP on stress. However, if the link persists, this has implications for the employers using PRP contracts, who may find that the economic benefits of higher productivity are lost due to stress-related work absences among PRP workers.

Supplementary material

Supplementary material is available at the *Oxford Economic Papers Journal* online which includes the data and statistical program that generates the results presented in the paper.

Funding

This work was supported by the UK Economic and Social Research Council [ES/R01163X/1].

Conflict of interest statement. None declared.

Acknowledgements

The authors would like to thank two anonymous referees and the editor for their helpful comments to improve the paper. The paper benefitted from the comments from seminar participants at the University of Aberdeen, and the CIPD Applied Research 2023 and Trier University Workshop on Performance Pay and Employee Outcomes 2022. Ethical approval was granted by the University of Aberdeen, School of Medicine, Medical Sciences and Nutrition Ethics Review Board (CERB/2021/3/2058). The usual disclaimer applies.

References

Allan, J. *et al.* (2021) 'Employment Contracts and Stress: Experimental Evidence', *Journal of Economic Behavior & Organization*, **187**: 360–73.

Allan, J. L., Bender, K. A., and Theodossiou, I. (2020) 'Performance Pay and Low-Grade Stress: An Experimental Study', *Work (Reading, Mass.)*, **67**: 449–57.

Andelic, N. *et al.* (2024) 'Performance-Related Pay, Mental and Physiological Health', *Industrial Relations: A Journal of Economy and Society*, **63**: 3–25.

Anderson, L. R., and Mellor, J. M. (2008) 'Predicting Health Behaviors with an Experimental Measure of Risk Preference', *Journal of Health Economics*, **27**: 1260–74.

Artz, B., Green, C. P., and Heywood, J. S. (2021) 'Does Performance Pay Increase Alcohol and Drug Use?', *Journal of Population Economics*, **34**: 969–1002.

Bandiera, O. *et al.* (2015) 'Matching Firms, Managers, and Incentives', *Journal of Labor Economics*, **33**: 623–81.

Bender, K. A., Green, C. P., and Heywood, J. S. (2012) 'Piece Rates and Workplace Injury: Does Survey Evidence Support Adam Smith?,' *Journal of Population Economics*, **25**: 569–90.

Bender, K. A., and Theodossiou, I. (2014) 'The unintended consequences of the rat race: the detrimental effects of performance pay on health', *Oxford Economic Papers*, **66**, 824–47.

Cable, D.M., and Judge, T.A. (1994) 'Pay Preferences and Job Search Decisions: A Person-Organization Fit Perspective', *Personnel Psychology*, **47**: 317–48.

Cadsby, C. B., Song, F., and Tapon, F. (2016) 'The Impact of Risk-Aversion and Stress on the Incentive Effect of Performance-Pay', in S. J. Georg and J. R. Hamman (eds) *Research in Experimental Economics*, pp. 189–227. Bingly, UK: Emerald Group Publishing Limited.

Cobb-Clark, D. A., Dahmann, S. C., and Kettlewell, N. (2022) 'Depression, Risk Preferences, and Risk-Taking behavior', *Journal of Human Resources*, **57**: 1566–604.

Cornelissen, T., Heywood, J. S., and Jirjahn, U. (2011) 'Performance Pay, Risk Attitudes and Job Satisfaction', *Labour Economics*, **18**: 229–39.

Curme, M., and Stefanec, N. (2007) 'Worker Quality and Labor Market Sorting', *Economics Letters*, **96**: 202–8.

Cutler, D. M., and Lleras-Muney, A. (2010) 'Understanding Differences in Health Behaviors by Education', *Journal of Health Economics*, **29**: 1–28.

Czeisler, C. A., and Klerman, E. B. (1999) 'Circadian and Sleep-Dependent Regulation of Hormone Release in Humans', *Recent Progress in Hormone Research*, **54**: 97–130.

Dahl, M. S., and Pierce, L. (2020) 'Pay-for-Performance and Employee Mental Health: Large Sample Evidence Using Employee Prescription Drug Usage', *Academy of Management Discoveries*, **6**: 12–38.

Dickerson, S. S., and Kemeny, M. E. (2004) 'Acute Stressors and Cortisol Responses: A Theoretical Integration and Synthesis of Laboratory Research', *Psychological Bulletin*, **130**: 355–91.

Dohmen, T., and Falk, A. (2011) 'Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender', *American Economic Review*, **101**: 556–90.

Dohmen, T. *et al.* (2011) 'Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences', *Journal of the European Economic Association*, **9**: 522–50.

Eriksson, T., and Villeval, M.C. (2008) 'Performance-Pay, Sorting and Social Motivation', *Journal of Economic Behavior & Organization*, **68**: 412–21.

Fischbacher, U. (2007) 'z-Tree: Zurich Toolbox for Ready-Made Economic Experiments', *Experimental Economics*, **10**: 171–8.

Freeman, R. B., and Kleiner, M. M. (2005) 'The Last American Shoe Manufacturers: Decreasing Productivity and Increasing Profits in the Shift from Piece Rates to Continuous Flow Production', *Industrial Relations: A Journal of Economy and Society*, **44**: 307–30.

Green, C., and Heywood, J. S. (2008) 'Does Performance Pay Increase Job Satisfaction?', *Economica*, **75**: 710–28.

Grund, C., and Sliwka, D. (2010) 'Evidence on Performance Pay and Risk Aversion', *Economics Letters*, **106**: 8–11.

Hruschka, D. J., Kohrt, B. A., and Worthman, C. M. (2005) 'Estimating Between- and Within-Individual Variation in Cortisol Levels Using Multilevel Models', *Psychoneuroendocrinology*, **30**: 698–714.

Kirschbaum, C., Pirke, K. M., and Hellhammer, D. H. (1993) 'The "Trier Social Stress test" – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting', *Neuropsychobiology*, **28**: 76–81.

Lazear, E.P. (2000) 'Performance Pay and Productivity', *American Economic Review*, **90**: 1346–61.

McCausland, W. D., Pouliakas, K., and Theodossiou, I. (2005) 'Some are Punished and Some are Rewarded: A Study of the Impact of Performance Pay on Job Satisfaction', *International Journal of Manpower*, **26**: 636–59.

McEwen, B. S. (1998) 'Stress, Adaptation, and Disease: Allostasis and Allostatic Load', *Annals of the New York Academy of Sciences*, **840**: 33–44.

Nagin, D. S. *et al.* (2002) 'Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment', *American Economic Review*, **92**: 850–73.

Nater, U.M. (2018) 'The Multidimensionality of Stress and its Assessment', *Brain, Behavior, and Immunity*, **73**: 159–60.

Nicolson, N. (2008) 'Measurement of Cortisol', in L. Luecken and L. Gallo (eds) *Handbook of Physiological Research Methods in Health Psychology*, pp. 37–74. Thousand Oaks, CA: Sage Publications.

Pruessner, J. C. *et al.* (2003) 'Two Formulas for Computation of the Area Under the Curve Represent Measures of total Hormone Concentration versus Time-Dependent Change', *Psychoneuroendocrinology*, **28**: 916–31.

Shearer, B., and Bellemare, C. (2006) 'Sorting, Incentives and Risk Preferences: Evidence from a Field Experiment', IZA Discussion Paper No. 2227.

Smith, A. (1937) *The Wealth of Nations*, New York: The Modern Library.

Starcke, K., and Brand, M. (2016) 'Effects of Stress on Decisions Under Uncertainty: A Meta-Analysis', *Psychological Bulletin*, **142**: 909–33.

Oxford Economic Papers, 2025, 00, 1–16
https://doi.org/10.1093/oep/gpaf008
Article