Brain and Cognition 102 (2016) 26–32

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Performance pressure and caffeine both affect cognitive performance, but likely through independent mechanisms

Julia J. Boere [a,1], Lizz Fellinger [a,1], Duncan J.H. Huizinga [a,1], Sebastiaan F. Wong [a,1], Erik Bijleveld [b,*]

[a] Department of Psychology, Utrecht University, The Netherlands
[b] Behavioural Science Institute, Radboud University, The Netherlands

ARTICLE INFO

Article history:
Received 21 April 2015
Revised 9 November 2015
Accepted 17 November 2015
Available online 10 December 2015

Keywords:
Performance pressure
Coffee
Modular arithmetic task
Dopamine
Eye Blink Rate (EBR)
BIS11

ABSTRACT

A prevalent combination in daily life, performance pressure and caffeine intake have both been shown to impact people's cognitive performance. Here, we examined the possibility that pressure and caffeine affect cognitive performance via a shared pathway. In an experiment, participants performed a modular arithmetic task. Performance pressure and caffeine intake were orthogonally manipulated. Findings indicated that pressure and caffeine both negatively impacted performance. However, (a) pressure vs. caffeine affected performance on different trial types, and (b) there was no hint of an interactive effect. So, though the evidence is indirect, findings suggest that pressure and caffeine shape performance via distinct mechanisms, rather than a shared one.

1. Introduction

In western society, people commonly hold the belief that coffee can be used to temporarily boost their cognitive performance. Accordingly, caffeine is the most widely used psychoactive stimulant in the world (Brunyé, Mahoney, Lieberman, & Taylor, 2010), and people are especially likely to ingest caffeine when they feel a strong need to perform to the best of their ability. Consider, for example, a magazine editor approaching a late-night deadline, a medical doctor getting ready to perform risky surgery, or a student on their way to their final exam. In high-stakes situations such as these, people may readily reach for a cup of coffee—and when it is finished, for another one.

Interestingly, the combination of performance pressure (e.g., due to highly valuable incentives) and caffeine intake may have important consequences for how well people perform. That is, as we will detail below, we expect that caffeine augments the (negative) effects of performance pressure on people's performance. To our knowledge, the combined effects of performance pressure and caffeine have not previously been studied, even though this combination is ubiquitous in daily life.

1.1. Performance pressure and performance

When people feel they are under high pressure to perform well, they have the tendency to perform below their ability (Beilock, Kulp, Holt, & Carr, 2004; DeCaro, Thomas, Albert, & Beilock, 2011; Eysenck & Calvo, 1992). This phenomenon is often referred to as *choking under pressure*, and has been shown to occur in various performance settings, including education, sports, and music. When such pressure-triggered drops in performance occur during cognitive-analytical tasks, these can well be accounted for by *distraction theory* (e.g., Ashcraft & Kirk, 2001; Beilock et al., 2004; Schmader, Johns, & Forbes, 2008). In essence, distraction theory suggests that when performance pressure increases, worries and thoughts regarding consequences of actions fill up working memory (WM). As a result, less WM is available for the main task at hand, causing performance to suffer. In support of distraction theory, performance decrements have been found to be especially pronounced on tasks (or trials) that rely heavily on WM (Beilock & Carr, 2005; Beilock et al., 2004).

While distraction theory provides a well-supported psychological account for choking under pressure on cognitive tasks, less is known about the biological mechanisms that may be involved. Nevertheless, previous research suggests that the mesolimbic and mesocortical dopamine pathways may play an important role in impairing performance when people are under pressure (Aarts et al., 2014; Arnsten, 2009; Bijleveld & Veling, 2014). When people are highly motivated to perform (which occurs under performance

* Corresponding author at: Behavioural Science Institute, Radboud University, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands.
E-mail address: e.bijleveld@psych.ru.nl (E. Bijleveld).
[1] The first four authors contributed equally to this article.

http://dx.doi.org/10.1016/j.bandc.2015.11.006
0278-2626/

pressure), higher order controlling PFC functions, such as WM and attention regulation, become impaired (Arnsten, 2009). During periods of high motivation, the release of dopamine (DA) is triggered in the prefrontal cortex (PFC), and through this mechanism the control of the PFC over other brain areas is modulated. At first, these increases in DA result in enhanced PFC control and performance (Pessoa & Engelmann, 2010). Importantly, however, the relationship between DA release and PFC control follows an inverted-U shape (Arnsten, 2009; Cools & Robbins, 2004). That is, when DA in the PFC rises above optimal levels, PFC control decreases again, leading to performance impairments on tasks that rely on the PFC. This mechanism is consistent with the idea that moderate motivational triggers (e.g., monetary rewards) tend to increase performance (Bijleveld, Custers, & Aarts, 2012; Garbers & Konradt, 2014; Krug & Braver, 2014), but also with the idea that extreme motivational triggers (e.g., intense performance pressure) have a detrimental effect (possibly accompanied by distracting thoughts, e.g., Beilock et al., 2004; Lee & Grafton, 2015).

As caffeine, like pressure, is known to increase DA levels (see below), the present research examines the possibility that caffeine ingestion makes people more prone to choke when under pressure. After all, when caffeine ingestion has already caused elevations of people's DA levels, subsequent performance pressure may arguably elevate DA levels in the PFC even further and cause the tipping point to be reached especially quickly, resulting in (more) pronounced drops in performance.

1.2. Caffeine and performance

The behavioral and neurochemical effects of caffeine (CA) have extensively been studied. It has been shown that CA, like performance pressure, affects both motor and cognitive-analytical performance (Brunyé et al., 2010; Lorist & Tops, 2003; Rogers, Heatherley, Mullings, & Smith, 2013; Smith, 2002). Moreover, similar to pressure, moderate doses of CA increase performance, while high doses (>500 mg) are thought to cause impairments (van der Stelt & Snel, 1998). On the psychological level, CA-induced increases in performance are thought to be mediated by a general increase in attention and processing speed, as well as a decrease in fatigue (Glade, 2010; Lorist & Tops, 2003; Smith, 2002). CA-induced decreases in task performance, as occur at higher doses, are often explained in terms of CA-triggered increases in anxiety and tension (Lorist & Tops, 2003; Smith, 2002).

On a neurochemical level, CA is thought to change performance through its effects on the endogenous neuromodulator adenosine. Specifically, caffeine readily crosses the blood–brain barrier and operates as an antagonist for adenosine A1 and A2A receptors (i.e., once CA binds to the A1 and A2A receptors it renders the binding of adenosine impossible). Adenosine A1 receptors are localized in almost all brain regions (in particular, in the hippocampus, cerebral cortex, cerebellar cortex, and thalamus), and are known to inhibit transmitter release in all types of neurons (Lorist & Tops, 2003). So while adenosine normally down-regulates other neurotransmitters (including DA), this function is impeded by CA. Via this route (i.e., by inhibiting adenosine functioning), CA may increase DA levels. In turn, as described earlier, DA affects performance by modulating the extent to which the PFC controls activity in other cortical brain areas (Arnsten, 2009). Taken together, the research addressed above suggests that performance pressure and CA may affect performance mediated by the same process: both may elevate DA levels in the PFC.

1.3. Coffee under pressure

The literature reviewed above suggests that performance pressure and caffeine may interact, with caffeine possibly strengthening the debilitating effect of performance pressure on cognitive-analytical performance. The present research was designed to test this idea. In the present experiment, participants were asked to solve a series of mathematical equations, presented in three blocks. The first of these blocks served as a pre-test measure of performance. The purpose of the second block was to stabilize performance. The purpose of the third block was to measure the effects of the pressure manipulation. That is, following Beilock et al. (2004, Experiment 1), half of the participants were subjected to performance pressure (vs. no performance pressure) during the third block. We expected that these participants would show a drop in performance from the first to the third block. Orthogonal to the pressure manipulation, half of the participants ingested 300 ml of caffeinated coffee (vs. decaffeinated coffee) before doing the task.

This design enabled us to examine the effect of caffeine independently of pressure (by looking at the pressure-free blocks, 1 and 2, as a function of caffeine intake). More importantly, it allowed us to test whether caffeine strengthened the effect of performance pressure. Specifically, in the pressure condition (but not in the no-pressure condition), we expect a drop in performance from blocks 1 to 3. In addition, we expect this drop to be steeper in people who have ingested caffeine (vs. people who have not).

To explore the biological processes that possibly play a role in boosting vs. debilitating performance, we also measured two known correlates of dopamine system functioning. First, we measured spontaneous Eye Blink Rate (EBR). Previous research suggests that differences in EBR baseline rates are correlated with striatal dopamine levels (Dreisbach et al., 2005; Karson, Dykman, & Paige, 1990). For example, in schizophrenics strongly heightened DA levels result in very high blink rates, whereas in Parkinson patients (very low DA availability) the opposite is observed (Swerdlow et al., 2003). This link was also found in experiments on monkeys, showing decreases in spontaneous blink rates after DA levels were depleted or receptors were blocked, while levels increased again after a DA agonist was administered (Taylor et al., 1999). Taking a pretest measure of spontaneous EBR allowed us to explore whether or not higher baseline blink rates predicted more pronounced pressure-induced drops in performance. To examine EBR changes over the course of the experiment (e.g., due to either of our manipulations), we measured EBR again near the end of the session. In addition, participants filled out the Baratt Impulsiveness Scale (BIS-11; Patton & Stanford, 1995; BIS-11 NL; Lijffijt & Barratt, 2005), a trait measure of impulsiveness that was previously found to correlate with DA activity in the midbrain (Buckholtz et al., 2010).

2. Method

2.1. Participants and design

One hundred and three participants (53 men, 50 women; age = 22.1), all students at Utrecht University, took part in the experiment. The experiment was reviewed and approved by the local ethical committee, and participants gave written informed consent. Before deciding to take part, they were told that the experiment would involve drinking coffee. Participants were randomly assigned to one condition of the 2(pressure: low vs. high) × 2(type of coffee: caffeine vs. decaf; double-blind) between-subjects design. Gauss' modular arithmetic task (Beilock et al., 2004) was used as a measure of analytical performance (see below).

Data from six participants were a priori excluded from analyses. Specifically, three participants were excluded because they did not drink the required amount of coffee. One further participant was excluded because the session was interrupted by a fire alarm. Another participant was excluded because this participant indicated afterward that he/she suffered from ADHD (which is

associated with abnormal dopamine functioning; del Campo, Chamberlain, Sahakian, & Robbins, 2011). Finally, one participant was excluded because performance on math accuracy was not above chance (<.55; following Beilock et al., 2004). These exclusions resulted in a final sample of 97 subjects.

2.2. Procedure

Upon arrival, participants were taken to a separate room, where the first EBR measurement was taken. After this measurement, they were sat down in a cubicle, which contained only a desk and a computer. When seated, they received the *caffeine manipulation*. Depending on condition, they received 300 ml of coffee that was caffeinated (approximately containing 170 mg caffeine; estimate based on Barone & Roberts, 1996) or decaffeinated. Then, the experimenter left, and participants filled out the Barratt Impulsiveness Scale (BIS-11, Patton & Stanford, 1995; BIS-11 NL, Lijffijt & Barratt, 2005) and some demographic questions on the computer. After filling out the questionnaire, they were instructed to finish the cup of coffee. Then, a neutral filler video started playing, to allow the CA to reach its full efficacy (after about 30 min; Lorist & Tops, 2003). Exactly 30 min after participants started filling out the questionnaires (which was, on average, after watching 20 min of video), participants were instructed to call the experimenter, who started the modular arithmetic task (see Measures). Then, participants performed the first two blocks of 24 trials.

After participants were instructed to again call the experimenter, who then (depending on the condition) delivered the *pressure manipulation*. Following Beilock et al. (2004), participants in the pressure condition were exposed to three concurrent, pressure-inducing manipulations. First, participants learned that the money they could earn for taking part in the experiment was contingent on their upcoming performance. Specifically, participants were told a performance score had been calculated over the previous block and that in order to earn a monetary reward (€8) their performance in the third block had to increase by 20% relative to this calculated performance score. They learned that if they would fail to improve, they would receive only €4. Second, participants were told that they were paired with another participant, who had ostensibly taken part earlier that day. They learned that both this previous participant (their "partner") and they themselves had to improve their performance, in order for either of them to receive the full monetary reward. However, they were also told that the previous participant had already been successful in increasing their performance. This way, participants were led to believe that both their own reward *and* the reward for their partner hinged on their performance during the upcoming block. Third, the experimenter set up a video camera on a tripod, announcing that the third block of trials would be recorded, for the purpose of later viewing by teachers and students interested in math performance. Taken together, participants in the high-pressure condition were exposed to several pressure-inducing triggers. By contrast, participants in the control condition were told to continue the task in the same way as before.

After the pressure manipulation, participants performed the third block of 24 modular arithmetic trials. After they were done, participants were again taken to a separate room for the second EBR measurement. Finally, they filled out a short questionnaire (to probe whether they experienced pressure, and to explore whether they noticed "something strange" about the coffee), were debriefed and paid €8, regardless of their performance during the experiment.

2.3. Task

The modular arithmetic task consists of three blocks of 24 modular arithmetic problems, such as $34 \equiv 18 \pmod 4$. Before starting the task, participants were explained that modular arithmetic problems can be solved by subtracting the second number from the first ($34 - 18 = 16$), and dividing the difference by the third number ($16/4 = 4$). If this division results in a whole number (like in this case, 4), the statement is *true*; otherwise, *false*. Participants indicated whether statements were true or false, by pressing the 'W' or 'O' keys. They were instructed to maximize speed and accuracy.

The problems were manipulated to be low, intermediate, or high in working memory demands. This was done by manipulating whether solving the problems required a single-digit no-borrow subtraction operation (low demand; e.g., $9 \equiv 2 \pmod 7$), a double-digit no-borrow subtraction operation (intermediate demand; e.g., $47 \equiv 15 \pmod 7$), or a double-digit borrow subtraction operation (high demand; e.g., $55 \equiv 27 \pmod 9$). Every block included 8 low-demand problems, 8 intermediate-demand problems, and 8 high-demand problems (Beilock et al., 2004). As such, participants needed approximately 2 min to complete each block. The statements within each block were presented in random order. Each *true* statement had a *false* counterpart within the same block (created by changing only the 'mod' number).

2.4. Eye Blink Rate (EBR) and Barratt Impulsiveness Scale (BIS)

EBR was measured twice: once before coffee intake, and once after the modular arithmetic task. Participants were seated in front of a Tobii T120 infrared eye tracker sampling at 120 Hz. After calibration, participants were asked to fixate on a black fixation cross for five minutes, while listening to calm music. From the resulting data, EBR was computed using a computer algorithm that counted the number of brief interruptions in the time-series data, which are indicative of eye blinks (for a description, see Aarts et al., 2012). The BIS-11 (Dutch translation; Lijffijt & Barratt, 2005) was completed via the computer, during coffee intake. Participants responded to 30 items (e.g., "I do things without thinking") by choosing between four response options (1–4, labeled as "seldom/never", "sometimes", "often", "almost always", respectively). Reliability of this scale was good, $\alpha = .84$.

3. Results

Precisely following Beilock et al. (2004), we started out by removing outliers from the data. The RT's for each math problem, and the means for each block were computed, for each participant individually. RT's that deviated more than 3SD from the relevant block mean were considered outliers and were discarded, along with their corresponding accuracy scores. This resulted in the removal of 70 trials (1.0%) in total.

3.1. Pressure

To test the hypothesis that pressure impairs performance (not yet considering caffeine), mean accuracy scores[2] were submitted to a 2(pressure: high vs. low, between subjects) × 2(block: 1 vs. 3, within subjects) × 2(problem demand: low vs. high, within subjects) ANOVA. This analysis revealed a main effect of problem demand,

[2] We chose to examine the effect of pressure in the full sample, rather than only in the decaf subsample, as this enabled us to do a more powerful statistical test due to the larger sample size. To provide transparency, examining the decaf condition only yields a highly similar pattern of results. That is, the same three-way interaction was present, $F(1,47) = 4.1$, $p = .050$, $\eta_p^2 = .08$. Like in the full sample, within low-demand problems, there were no significant changes in performance, $Fs(1,47) < 1.4$, $ps > .256$, $\eta_p^2 s < .03$. Within high-demand problems, there was a performance increase in the no-pressure condition, $F(1,47) = 4.5$, $p = .039$, $\eta_p^2 = .09$, and a performance decrease in the high-pressure condition, $F(1,47) = 3.4$, $p = .071$, $\eta_p^2 = .07$ (though only marginally significant; presumably due to loss of power).

$F(1,95) = 95.5$, $p < .001$, $\eta_p^2 = .50$, indicating that participants were less accurate on high-demand problems. Also, there was a pressure × block interaction, $F(1,95) = 12.2$, $p = .001$, $\eta_p^2 = .11$. Importantly, these effects were qualified by the predicted pressure × block × problem demand interaction, $F(1,95) = 7.6$, $p = .007$, $\eta_p^2 = .07$. To interpret this pattern of results, we inspected the pattern of means (Fig. 1).

Findings essentially replicated previous work (Beilock et al., 2004, Experiment 1). On low-demand problems, accuracy increased from block 1 to block 3. However, this increase was not significant (low-pressure condition, $F[1,95] = .1$, $p = .739$, $\eta_p^2 < .01$; high-pressure condition, $F[1,95] = .6$, $p = .453$, $\eta_p^2 < .01$). On high-demand problems, accuracy significantly increased from block 1 to block 3, but only in the low-pressure condition, $F(1,95) = 5.4$, $p = .022$, $\eta_p^2 = .05$. By contrast, in the high pressure condition, there was a decrease in accuracy, $F(1,95) = 9.8$, $p = .002$, $\eta_p^2 = .09$. This pattern of findings indicates that pressure harms performance on mathematical problem solving, but only on high-demand problems.

Following previous work (Beilock et al., 2004), we next examined whether these results were not an artefact of a speed–accuracy tradeoff (e.g., when facing high-demand problems under pressure, people may decide to act faster, at the expense of accuracy). To do so, we submitted the RT's (of trials in which people were accurate) to the three-factor ANOVA. This analyses revealed a main effect of block, $F(1,95) = 48.0$, $p < .001$, $\eta_p^2 = .34$, indicating people were faster in block 3 (vs. block 1). There was a main effect of problem demand, $F(1,95) = 424.8$, $p < .001$, $\eta_p^2 = .82$, indicating that people were faster on low-demand problems when compared to high-demand problems. These effects were qualified by the block × demand interaction, $F(1,95) = 4.0$, $p = .048$, $\eta_p^2 = .04$, suggesting that people became faster in block 3 especially for high-demand problems. All other effects were not significant, Fs < 3.7, p's > .059, $\eta_p^2 < .4$, including the pressure × block × problem demand three way-interaction, $F(1,95) = 1.2$, $p = .268$, $\eta_p^2 = .01$. This indicates that the specific drop in accuracy for high-demand problems, triggered by pressure, cannot be explained by changes in speed–accuracy tradeoffs.

To more closely examine the possibility of a speed-accuracy trade-off, we conducted a 2(pressure: high vs. low) × 2(block: 1 vs. 3) ANOVA on the accuracy on high-demand problems, while controlling for differences in RT as they occurred from block 1 to block 3 (i.e., block 1–block 3). Replicating previous work (Beilock et al., 2004), the predicted pressure × block interaction was significant, $F(1,94) = 19.4$, $p < .001$, $\eta_p^2 = .17$. This analysis substantiates the idea that the pressure-induced drop in accuracy on high-demand problems cannot be explained by a speed–accuracy tradeoff.

3.2. Caffeine

To examine the effects of caffeine on performance (not yet regarding pressure), we analyzed people's performance averaged over the blocks before the pressure manipulation (i.e., blocks 1 and 2) using a 2(demand: low vs. high, within subjects) × 2(type of coffee: caffeine vs. decaf, between subjects) ANOVA. This analysis yielded a main effect of demand, $F(1,95) = 77.2$, $p < .001$, $\eta_p^2 = .45$, indicating lower accuracy on high-demand problems. However, there was neither a main effect of type of coffee, $F(1,95) = .8$, $p = .386$, $\eta_p^2 < .01$, nor a type of coffee × demand interaction, $F(1,95) = .2$, $p = .645$, $\eta_p^2 < .01$. So, there is no evidence that our caffeine manipulation directly impacted math performance.

To examine the nature of this null effect, we considered the possibility that the administered caffeine dose was too low to exert an effect. If this were true, we reasoned, the manipulation may still have had an effect on people who already ingested caffeine in the hours preceding the experiment. After all, caffeine has a half life of 3–5 h (Lorist & Tops, 3002), so it may have been the case that some participants were already under the influence of caffeine when the session began. We conducted a 2(demand: low vs. high, within subjects) × 2(type of coffee: caffeine vs. decaf, between subjects) × 2(history: ingested coffee in the past four hours vs. no coffee in the past four hours) ANOVA.[3] This analysis revealed a main effect of demand, $F(1,91) = 80.9$, $p < .001$, $\eta_p^2 = .47$, indicating that accuracy was lower for high-demand problems. There was a main effect of history, $F(1,91) = 8.8$, $p = .004$, $\eta_p^2 = .09$, indicating that people who had recently ingested coffee performed worse than people who had not. Importantly, this effect was qualified by the type of coffee × history interaction, $F(1,91) = 8.1$, $p = .005$, $\eta_p^2 = .08$, indicating that caffeine manipulation negatively impacted performance in people who had coffee before the experiment, $F(1,91) = 8.5$, $p = .004$, $\eta_p^2 = .09$, but not in people who had not, $F(1,91) = .6$, $p = .448$, $\eta_p^2 < .01$. However, this interaction was not qualified by the three-way interaction, $F(1,91) = 1.1$, $p = .296$, $\eta_p^2 = .01$, suggesting that caffeine affected low-demand and high-demand problems to the same extent. The pattern of means is plotted in Fig. 2.

We conducted the same ANOVA to examine RT's. The only effect that was significant was the main effect of demand, $F(1,91) = 275.3$, $p < .001$, $\eta_p^2 = .75$, indicating that people were slower on high-demand trials. Hence, there is no reason to assume that the specific pattern of results found for accuracy can be explained by a speed–accuracy tradeoff.

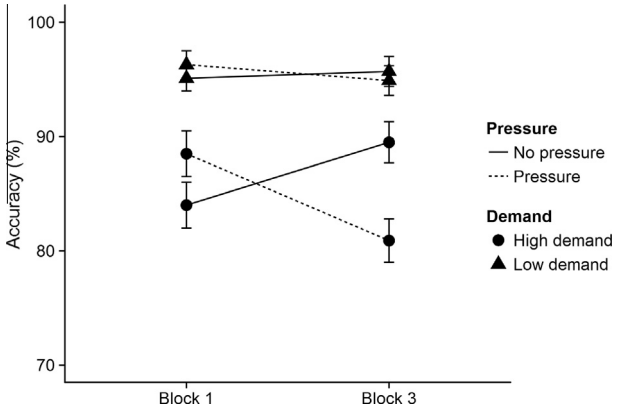

Fig. 1. The effects of performance pressure and problem demand on accuracy. Error bars indicate standard errors.

3.3. Pressure and caffeine

Next, we tested whether the pressure-induced drop in performance on high-demand trials was affected by caffeine ingestion. To that end, we computed an accuracy change score (i.e., block 3–block 1) and we submitted this score to a 2(type of coffee: caffeine vs. decaf) × 2(pressure: low vs. high) ANOVA. This analysis yielded the expected effect of pressure, $F(1,93) = 15.1$, $p < .001$, $\eta_p^2 = .14$, reflecting the finding that pressure harmed performance. However, there was neither a main effect of type of coffee, $F(1,93) = .9$, $p = .347$, $\eta_p^2 = .01$, nor a type of coffee × pressure interaction, $F(1,93) < .1$, $p = .884$, $\eta_p^2 < .01$. So, there is no direct evidence for the hypothesis that caffeine strengthens the performance-debilitating effect of pressure.

Next, in parallel with the accuracy scores, we examined changes in RT's for the high-demand trials (block 3–block 1) with a 2(type of coffee: caffeine vs. decaf) × 2(pressure: low vs. high) ANOVA.

[3] As two participants did not indicate when they last had coffee, this analysis was performed with $N = 95$.

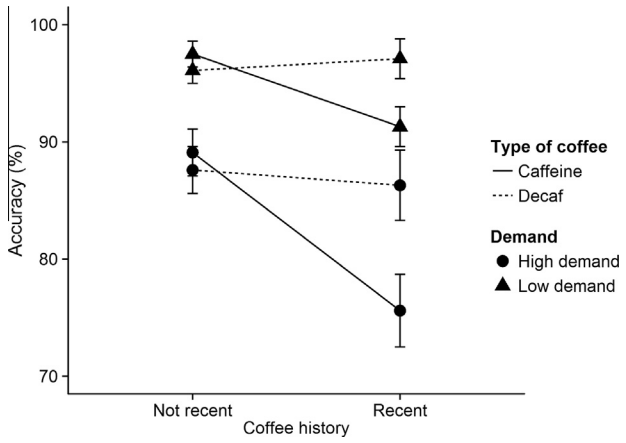

Fig. 2. The effects of coffee history (i.e., coffee before the experiment) and caffeine ingestion (during the experiment) on accuracy. Error bars indicate standard errors.

This analysis yielded no significant effects, Fs $(1,93) < 2.5$, ps $> .120$, η_p^2's $< .03$.

Given that the main effect of type of coffee became apparent only after considering previous coffee intake, a logical next step would be to examine the type of coffee × pressure interaction separately for people who had coffee before the experiment (vs. not). To take the same analytic strategy as before, however, this would require us to conduct a 2 (pressure) × 2(type of coffee) × 2(coffee history, quasi-experimental) all-between-subjects ANOVA. As cell sizes would be too small (range: 3–19) to draw sensible conclusions from this analysis, we chose to refrain from conducting it. We did, however, explore this possibility further using a Bayesian approach. For the present purposes, a key advantage of this approach is that it can be used to draw clearer conclusions about the meaning of null effects (e.g., Dienes, 2014; Kass & Raftery, 1995; Kruschke, 2011). Specifically, in a regular ANOVA, a null effect can mean that the null hypothesis is more likely than the alternative hypothesis. However, a null effect can also mean that more evidence is needed to adjudicate between the null and the alternative (Dienes, 2014). The Bayesian analyses presented below allow for a dissociation of these two possibilities, thus allowing us to draw more precise conclusions from our data.

3.4. Pressure and caffeine: Bayes Factors

In a Bayesian approach to hypothesis testing, analyses do not just evaluate the extremeness of the data under the null hypothesis; rather, they evaluate the likelihood of the data *both* under the null hypothesis and under the alternative hypothesis. The ratio of these two likelihoods, called the Bayes Factor (BF), reflects which of the two hypotheses is more likely, and to what extent (Kass & Raftery, 1995; Kruschke, 2011; Rouder, Speckman, Sun, Morey, & Iverson, 2009). A BF of 5, for example, means that the alternative is three times more likely than the null, given the data; conversely, a BF of 1/5 means that the null is five times more likely than the alternative. BF's of 3 and higher (and 1/3 and lower) are commonly considered to be meaningful (Kass & Raftery, 1995).

We re-examined our main prediction by conducting a Bayesian ANOVA on accuracy change scores (i.e., block 3–block 1, difficult trials only), with pressure and type of coffee as independent variables. BF's were computed using BayesFactor software (Morey & Rouder, 2015), which implements a Bayesian hypothesis testing algorithm with an established set of default settings (specifically, Bayesian ANOVA using JZS priors; Rouder, Morey, Speckman, & Province, 2012). The main results from this analysis are presented in Table 1. Inspection of Table 1 reveals that the model with pressure as the only predictor (model 4), is clearly the preferred one. Given the data, this pressure-only model is 11.0 times[4] as likely to be true compared to the model that includes the pressure × type of coffee interaction (model 2). So, speaking against our main hypothesis, there is positive evidence (Kass & Raftery, 1995) against the presence of the predicted pressure × type of coffee interaction in the whole sample.

Then, we examined the possibility that type of coffee may have modulated the effect of pressure only among people who had coffee before the experiment. After all, our exploratory analyses presented in Section 3.2 suggested that only this subgroup of people was sensitive to the effects of type of coffee. We conducted the same pressure × type of coffee Bayesian ANOVA as before, but now separately for people who previously had ingested coffee vs. people who had not. The results from this analysis are reported in the two rightmost columns of Table 1. Though the evidence is weaker within both two subgroups compared to the entire sample (presumably, due to the fact that there is simply less data in the subgroups), the pattern of results in both groups was very similar to the pattern of results in the full sample. Specifically, the model with pressure as the only predictor (model 4) was clearly preferred over all other models. In both subgroups, given the data, this model was more likely to be true—2.7 and 3.2 times, respectively—than the model that included the pressure × type of coffee interaction (model 2). So, again speaking against our main hypothesis, there is some evidence (though "not worth more than a bare mention"; Kass & Raftery, 1995) against the existence of a pressure × type of coffee interaction among people who had coffee before the experiment.

3.5. EBR and BIS

Finally, we were interested to explore (a) whether caffeine increased EBR from before to after the experimental session and (b) whether BIS scores and baseline EBR were related to pressure-induced drops in performance. To test for caffeine-triggered changes in EBR, we conducted a 2(type of coffee: caffeine vs. decaf) × 2(time: t_0 vs. t_1) ANOVA. Although EBR unexpectedly decreased from t_0 to t_1, $F(1,90) = 15.2$, $p < .001$, $\eta_p^2 = .14$, there was no main effect of type of coffee, $F(1,90) = .3$, $p = .578$, $\eta_p^2 < .1$, and no type of coffee × time interaction, $F(1,90) < .1$, $p = .978$, $\eta_p^2 < .01$. To examine whether the pressure–performance relation was moderated by BIS, we carried out a General Linear Model (GLM) analysis, in which we predicted changes in accuracy on high-demand trials (block 3–block 1) from pressure (low vs. high, between subjects) and BIS (continuous, between subjects). The effect of pressure was not moderated by BIS, $F(1,93) = .7$, $p = 399$, $\eta_p^2 < .01$. We carried out the same GLM analysis, but now with EBR at t_0 as a predictor. The effect of pressure was not moderated by EBR, $F(1,90) = 1.3$, $p = .259$, $\eta_p^2 = .01$. We will return to these results in the discussion.

4. Discussion

In this research, we set out to investigate whether performance pressure and caffeine intake affect cognitive performance via a shared biological route. First, independently of caffeine, we examined how performance was affected by performance pressure. After exposure to pressure, people performed worse than they did before. Importantly, no such decline in performance occurred in people who were *not* exposed to pressure. In line with distraction

[4] The BF of 11.0 was computed by dividing the BF of model 4 by the BF of model 2. This was done to directly compare the two models against each other, rather than to compare both models against an intercept-only model (as in Table 1).

Table 1
Bayesian analysis of the effects of pressure and type of coffee on accuracy change scores (block 3 – block 1) during difficult trials.

	Predictors in model	Bayes Factor		
		All participants ($N = 97$)	Coffee before experiment ($N = 29$)	No coffee before experiment ($N = 66$)
1	Type of coffee	0.3	1.8	1.2
2	Pressure + type of coffee + pressure × type of coffee	10.9	2.6	2.2
3	Pressure + type of coffee	37.2	4.7	5.8
4	Pressure	119.8	6.9	7.0

Note. All Bayes Factors (BFs) are relative to a model that includes only an intercept. Models are sorted according to their likelihood given the data; in all subsamples, model 1 was least likely, model 4 was most likely.

theory, replicating previous work (Beilock et al., 2004), pressure-induced drops in performance specifically occurred on math problems that required double-digit carry operations (i.e., high-WM-demanding problems), but not on easier problems. Second, independently of performance pressure, we examined how performance was affected by caffeine. On the group level, caffeine did not have any reliable effects. However, in people who had already consumed caffeine in the four hours preceding the session, caffeine ingestion sharply decreased performance (for a discussion of the dose–response relation between caffeine and performance see Lorist & Tops, 2003). Importantly, unlike the effect of pressure, the caffeine-induced performance decrement was *not* specific for high-demand math problems—instead, the effect occurred to a similar extent on low-demand and high-demand trials.

To test the idea that caffeine enhances the performance-degrading effect of pressure, we examined the combined effects of pressure and caffeine. We found evidence against such a modulation. It is important to note, though, that this null effect may be explained by the absence of a strong group-level effect of caffeine. After all, caffeine affected performance only in participants who had already consumed coffee before the experiment. Therefore, we further explored whether a pressure × caffeine interaction may have been present only among people who had already ingested caffeine. Based on our analysis of Bayes Factors, this possibility seemed unlikely to be true.

Taken together, the pattern of findings speaks against the idea that pressure and caffeine influence performance via a common biological pathway. First, pressure and caffeine affected performance on different types of trials. Second, although both pressure and caffeine independently affected performance, there was no hint of an interactive effect. Our findings, then, are consistent with the idea that the behavioral effects of pressure and caffeine are due to independent mechanisms. It may still be the case that both these mechanisms involve DA transmission, as the underlying neural infrastructure is highly complex. However, there are also alternative candidate mechanisms, which we will now discuss.

First, non-DA-related mechanisms may account for pressure-induced performance decrements. Norepinephrine (NE) is an important candidate, since NE strongly modulates PFC functioning (Arnsten, 2009). Even though NE is synthesized from DA, their levels do not necessarily run parallel (LeBlanc & Ducharme, 2007) and they may therefore serve different functions. Interestingly, like DA, NE levels also affect performance according to an inverted U-curve (Aston-Jones & Cohen, 2005; Arnsten, 2009). The inverted U-shaped relationship of NE release and performance is thought to be caused by differential engagement of adrenergic receptors: smaller levels increase alertness through their engagement of α2-adrenergic receptors; higher levels engage α1-adrenergic and β-adrenergic receptors, impairing PFC functioning (Arnsten, 2009). Besides NE, glutamatergic transmission is another candidate mechanism. Acute stressors enhance glutamatergic transmission, and also here, glutamate levels have a certain optimum level that can be exceeded (Yuen et al., 2009). So, the effects of pressure that emerged in this study may have been mediated by NE or glutamate, rather than by the hypothesized DA modulation.

Similarly, non-DA-related mechanisms may also account for the effect of caffeine on performance. As caffeine is an adenosine antagonist, it inhibits downregulation not only of DA, but also of NE and glutamate. So, both performance pressure and caffeine may affect several neurotransmitter systems that are relevant to cognitive functioning (Brunyé et al., 2010; Carli & Invernizzi, 2014; Lorist & Tops, 2003), suggesting intriguing avenues for future research.

We expected performance pressure during the experiment to push participants' dopamine levels beyond their optimum. In interaction with caffeine intake, this was expected to result in even worse performance. However, no such extra performance decrement appeared. A possible explanation for this null effect is that the time between caffeine intake and testing in our experiment was 30 min, whereas the estimated peak of caffeine in blood plasma levels is 30–60 min post intake (Lorist & Tops, 2003). It may have been the case that the main task was performed too soon after CA intake for the relevant neurochemical processes to fully unfold. In support of this reasoning, participants who had ingested coffee in the four hours prior to the experiment (vs. participants who had not) performed worse, independently of experimental condition.

We manipulated performance pressure with a procedure that consisted of three components (i.e., the prospect of a monetary reward, ostensible expectations by other people, being recorded; Beilock et al., 2004). An important advantage of this manipulation is that it combines various sources of pressure that together produce a replicable effect on performance. It should be mentioned, though, that this procedure cannot disentangle the effects of the three sources. For example, our study does not speak to the individual effects of monetary reward vs. being recorded by a video camera. Examining each of these sources of pressure in isolation constitutes an important direction for future research, as they may well have diverging psychological consequences (DeCaro et al., 2011; Mesagno, Harvey, & Janelle, 2011).

The EBR and BIS measures we took in the present study were hypothesized to predict individual differences in the strength of the effects of caffeine and pressure. However, neither EBR (before the experiment) nor BIS-scores predicted responses to pressure or to caffeine. These null results can be explained in three ways. First, the DA system is not at all involved in mediating the effects of pressure and caffeine. This is possible, but unlikely based on prior work (Arnsten, 2009; Lorist & Tops, 2003). Second, the DA system is involved, but baseline DA is not so important. Again, this is possible but also unlikely given previous findings (Cools & D'Esposito, 2011). Third, although EBR and BIS correlate with aspects of DA transmission, it may be the case that this correlation is too weak to have any predictive value, at least in the domain of high-stakes performance. This last argument could be true. Although EBR and BIS both reflect properties of the DA system (Aarts et al., 2012; Buckholtz et al., 2010; Taylor et al., 1999), they are clearly rather indirect indicators. Future research (e.g., with pharmacological manipulations) is needed to more closely evaluate the usefulness of putative behavioral and self-report markers of baseline DA.

When comparing EBR measurements before and after the experiment, an unexpected, interesting decrease in EBR was found.

There were are some aspects of the experimental procedure that might have caused this decrease. First, upon completion of the main experimental task, informal reports indicated that participants generally expressed relaxation, sometimes even relief. Second, during the second EBR measurement, participants often expressed boredom (as they had to look at a fixation cross for five minutes again). So, it could be the case that participants were generally less aroused (or otherwise affected by time-on-task) at the end of the session, which could explain the decrease in EBR (Barbato et al., 2000).

4.1. Conclusion

We examined the notion that performance pressure and caffeine affect performance via a common mechanism, but we found no evidence for this idea. Instead, pressure and caffeine both affected performance, but this likely occurred via different routes. While the exact nature of the underlying biological processes is yet to be delineated, the present research suggests that having a cup of coffee (and another one) before a high-stakes performance situation is unlikely to strongly augment the effects of performance pressure.

References

Aarts, H., Bijleveld, E., Custers, R., Dogge, M., Deelder, M., Schutter, D., & van Haren, N. E. M. (2012). Positive priming and intentional binding: Eye-blink rate predicts reward information effects on the sense of agency. *Social Neuroscience, 7*, 105–112. http://dx.doi.org/10.1080/17470919.2011.590602.

Aarts, E., Wallace, D. L., Dang, L. C., Jagust, W. J., Cools, R., & D'Esposito, M. (2014). Dopamine and the cognitive downside of a promised bonus. *Psychological Science, 25*, 1003–1009. http://dx.doi.org/10.1177/0956797613517240.

Arnsten, A. F. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience, 10*, 410–422. http://dx.doi.org/10.1038/nm2648.

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130*, 224–237. http://dx.doi.org/10.1037/0096-3445.130.2.224.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28*, 403–450. http://dx.doi.org/10.1146/annurev.neuro.28.061604.135709.

Barbato, G., Ficca, G., Muscetolla, G., Fichelle, M., Beatrice, M., & Rinaldi, F. (2000). Diurnal variation in spontaneous eyeblink-rate. *Psychiatry Research, 93*, 145–151. http://dx.doi.org/10.1016/S0165-1781(00)00108-6.

Barone, J. J., & Roberts, H. R. (1996). Caffeine consumption. *Food and Chemical Toxicology, 34*, 119–129. http://dx.doi.org/10.1016/0278-6915(95)00093-3.

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Science, 16*, 101–105. http://dx.doi.org/10.1111/j.0956-7976.2005.00789.x.

Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General, 133*, 584–600. http://dx.doi.org/10.1037/0096-3445.133.4.584.

Bijleveld, E., Custers, R., & Aarts, H. (2012). Adaptive reward pursuit: How effort requirements affect unconscious reward responses and conscious reward decisions. *Journal of Experimental Psychology: General, 141*, 728–742. http://dx.doi.org/10.1037/a0027615.

Bijleveld, E., & Veling, H. (2014). Separating chokers from non-chokers: Predicting real-life tennis performance under pressure from behavioral tasks that tap into working memory functioning. *Journal of Sport and Exercise Psychology, 36*, 347–356. http://dx.doi.org/10.1123/jsep.2013-0051.

Brunyé, T. T., Mahoney, C. R., Lieberman, H. R., & Taylor, H. A. (2010). Caffeine modulates attention network function. *Brain and Cognition, 72*, 181–188. http://dx.doi.org/10.1016/j.bandc.2009.07.013.

Buckholtz, J. W., Treadway, M. T., Cowan, R. L., Woodward, N. D., Li, R., Ansari, M. S., ... Zald, D. H. (2010). Dopaminergic network differences in human impulsivity. *Science, 329*, 532. http://dx.doi.org/10.1126/science.1185778.

Carli, M., & Invernizzi, R. W. (2014). Serotoninergic and dopaminergic modulation of cortico-striatal circuit in executive and attention deficits induced by NMDA receptor hypofunction in the 5-choice serial reaction time task. *Frontiers in Neural Circuits, 8*, 1–28. http://dx.doi.org/10.3389/fncir.2014.00058.

Cools, R., & D'Esposito, M. (2011). Inverted-U-shaped dopamine actions on human working memory and cognitive control. *Biological Psychiatry, 69*, e113–e125. http://dx.doi.org/10.1016/j.biopsych.2011.03.028.

Cools, R., & Robbins, T. W. (2004). Chemistry of the adaptive mind. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 362*, 2871–2888. http://dx.doi.org/10.1098/rsta.2004.1468.

DeCaro, M. S., Thomas, R. D., Albert, N. B., & Beilock, S. L. (2011). Choking under pressure: Multiple routes to skill failure. *Journal of Experimental Psychology: General, 140*, 390–406. http://dx.doi.org/10.1037/a0023466.

del Campo, N., Chamberlain, S. R., Sahakian, B. J., & Robbins, T. W. (2011). The roles of dopamine and noradrenaline in the pathophysiology and treatment of attention-deficit/hyperactivity disorder. *Biological Psychiatry, 69*, e145–e157. http://dx.doi.org/10.1016/j.biopsych.2011.02.036.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*, 781. http://dx.doi.org/10.3389/fpsyg.2014.00781.

Dreisbach, G., Müller, J., Goschke, T., Strobel, A., Schulze, K., Lesch, K. P., & Brocke, B. (2005). Dopamine and cognitive control: The influence of spontaneous eyeblink rate and dopamine gene polymorphisms on perseveration and distractibility. *Behavioral Neuroscience, 119*, 483–490. http://dx.doi.org/10.1037/0735-7044.119.2.483.

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion, 6*, 409–434. http://dx.doi.org/10.1080/02699939208409696.

Garbers, Y., & Konradt, U. (2014). The effect of financial incentives on performance: A quantitative review of individual and team-based financial incentives. *Journal of Occupational and Organizational Psychology, 87*, 102–137. http://dx.doi.org/10.1111/joop.12039.

Glade, M. J. (2010). Caffeine – Not just a stimulant. *Nutrition, 26*, 932–938. http://dx.doi.org/10.1016/j.nut.2010.08.004.

Karson, C. N., Dykman, R. A., & Paige, S. R. (1990). Blink rates in schizophrenia. *Schizophrenia Bulletin, 16*, 345–354. http://dx.doi.org/10.1093/schbul/16.2.345.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773–795. http://dx.doi.org/10.1080/01621459.1995.10476572.

Krug, M. K., & Braver, T. S. (2014). Motivation and cognitive control: Going beyond monetary incentives. In E. Bijleveld & H. Aarts (Eds.), *The psychological science of money* (pp. 137–162). New York: Springer. http://dx.doi.org/10.1007/978-1-4939-0959-9_7.

Kruschke, J. K. (2011). *Doing Bayesian data analysis: A tutorial with R and BUGS.* Waltham, MA: Academic Press.

LeBlanc, J., & Ducharme, M. B. (2007). Plasma dopamine and noradrenaline variations in response to stress. *Physiology and Behavior, 91*, 208–211. http://dx.doi.org/10.1016/j.physbeh.2007.02.011.

Lee, T. G., & Grafton, S. T. (2015). Out of control: Diminished prefrontal activity coincides with impaired motor performance due to choking under pressure. *NeuroImage, 105*, 145–155. http://dx.doi.org/10.1016/j.neuroimage.2014.10.058.

Lijffijt, M., & Barratt, E. S. (2005). Barratt impulsiveness scale 11 – Dutch version. Retrieved from <http://www.impulsivity.org/measurement/bis11_Dutch>.

Lorist, M. M., & Tops, M. (2003). Caffeine, fatigue, and cognition. *Brain and Cognition, 53*, 82–94. http://dx.doi.org/10.1016/S0278-2626(03)00206-9.

Mesagno, C., Harvey, J. T., & Janelle, C. M. (2011). Self-presentation origins of choking: Evidence from separate pressure manipulations. *Journal of Sport and Exercise Psychology, 33*, 441–459.

Morey, R. D., & Rouder, J. N. (2015). BayesFactor [Computer Software; version 0.9.12-2]. Retrieved from <http://bayesfactor.blogspot.nl/>.

Patton, J. H., & Stanford, M. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology, 51*, 768–774.

Pessoa, L., & Engelmann, J. B. (2010). Embedding reward signals into perception and cognition. *Frontiers in Neuroscience, 4*, 17. http://dx.doi.org/10.3389/fnins.2010.00017.

Rogers, P. J., Heatherley, S. V., Mullings, E. L., & Smith, J. E. (2013). Faster but not smarter: Effects of caffeine and caffeine withdrawal on alertness and performance. *Psychopharmacology, 226*, 229–240. http://dx.doi.org/10.1007/s00213-012-2889-4.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*, 356–374. http://dx.doi.org/10.1016/j.jmp.2012.08.001.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review, 16*, 225–237. http://dx.doi.org/10.3758/PBR.16.2.225.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*, 336–356. http://dx.doi.org/10.1037/0033-295X.115.2.336.

Smith, A. (2002). Effects of caffeine on human behavior. *Food and Chemical Toxicology, 40*, 1243–1255. http://dx.doi.org/10.1016/S0278-6915(02)00096-0.

Swerdlow, N. R., Wasserman, L. C., Talledo, J. A., Casas, R., Bruins, P., & Stephany, N. L. (2003). Prestimulus modification of the startle reflex: Relationship to personality and psychological markers of dopamine function. *Biological Psychology, 62*, 17–27. http://dx.doi.org/10.1016/S0301-0511(02)00090-X.

Taylor, J. R., Elsworth, J. D., Lawrence, M. S., Sladek, J. R., Jr., Roth, R. H., & Redmond, D. E. Jr., (1999). Spontaneous blink rates correlate with dopamine levels in the caudate nucleus of MPTP-treated monkeys. *Experimental Neurology, 158*, 214–220. http://dx.doi.org/10.1006/exnr.1999.7093.

van der Stelt, O., & Snel, J. (1998). Caffeine and human performance. In J. Snel & M. M. Lorist (Eds.), *Nicotine, caffeine and social drinking* (pp. 167–183). Amsterdam, The Netherlands: Harwood.

Yuen, E. Y., Liu, W., Karatsoreos, I. N., Feng, J., McEwen, B. S., & Yan, Z. (2009). Acute stress enhances glutamatergic transmission in prefrontal cortex and facilitates working memory. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 14075–14079. http://dx.doi.org/10.1073/pnas.0906791106.