Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Assessing Decision-Making Accuracy, Error Rates, and Guessing Propensity Under Dietary Caffeine Doses: A Within-Subject Pilot Investigation

Yong Sheng Raymond Teoh
University Teknologi Malaysia
Rubita Sudirman
rubita@fke.utm.my

University Teknologi Malaysia
Eileen Lee Ming Su
University Teknologi Malaysia

Research Article

Keywords: Caffeine, Decision Making, Impulsive Behavior, Reaction Time, Mental Health

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11474980/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Introduction:

Mental health includes how people think and make judgements, not only the absence of illness. Caffeine is believed to leave drinkers "on edge" and quick to act before thinking, yet whether everyday doses make decisions hasty or error-prone has rarely been tested.

Methods

Nine enrolments of eight young adults each completed three single-blind sessions in counterbalanced order, drinking placebo, 70 mg or 140 mg of caffeine (0.02 to 2.98 mg/kg). Flanker, visual short-term memory and serial-subtraction tasks and mood ratings were completed before and 5, 30 and 55 min after the drink, with skin conductance and heart rate recorded. Errors, guessing, response bias and time spent checking answers were compared with placebo, adjusting for each session's pre-dose score. Equivalence tests asked whether a rash effect of half a standard deviation could be ruled out.

Results

Very fast responses were rare (11 of 4,559 under 300 ms). After the drink, flanker error rates were 1.6% with placebo, 0.8% with 70 mg and 0.7% with 140 mg. At both doses, a rise of the predefined size in flanker or arithmetic errors, or a shift towards guessing on the memory task, was ruled out ($p \leq 0.035$). Errors were not fast guesses, and answer checking did not decline. Jitteriness, anxiety and stress did not rise significantly, and feeling jittery was unrelated to how participants responded.

Conclusion

At everyday doses, caffeine did not make decisions more hasty or error-prone. Feeling on edge and acting rashly appeared to be separate.

Introduction

Mental health is more than the absence of mental illness. It also covers everyday mental functioning: how well people pay attention, weigh information and make sound judgements [1]. Caffeine is the substance most often used to change that functioning on purpose. People drink coffee or tea to feel more alert, to think faster and to keep going through long days [2, 3]. Even at doses far below those that cause clinical anxiety, then, caffeine shapes the mental state in which ordinary decisions are made [4].

Caffeine also has a reputation for an unwanted side. Drinkers describe feeling jittery, tense or "on edge", especially after larger doses [5]. It is a short step from that feeling to the belief that caffeine also makes

people rash: quick to answer before they have properly looked, more inclined to guess, and so more prone to mistakes. Arousal theory points the same way. If performance rises with arousal only up to a point and then falls [6], a stimulant that pushes arousal past that point should harm the quality of decisions, and caffeine's effect on performance has long been known to depend on how aroused and how impulsive a person is to begin with [7].

Rashness can take two forms [8]. Impulsive action is responding too quickly or failing to hold back a response. Impulsive choice is preferring a risky or immediate reward over a safer or better one. A related idea is the trade-off between speed and accuracy: a person can answer faster by accepting more errors, and a rash decision-maker is one who has moved too far toward speed [9].

The evidence that caffeine produces either form of rashness is mixed. Surveys find that people who consume a lot of caffeine also score higher on sensation seeking and impulsivity [10], and in young adults with gambling problems, higher habitual intake went with more impulsive personality scores and poorer betting decisions [11]. Caffeine use has also been linked with aggression and impulsiveness in adults with and without aggression problems [12]. These studies, however, show who chooses to drink a lot of caffeine, not what a cup of coffee does to a decision. Experiments that give caffeine directly tell a different story. A dose of 3 mg/kg did not weaken the ability to hold back a response on flanker and stop-signal tasks [13]. In healthy adults kept awake for 75 hours, caffeine neither worsened nor rescued decision-making on a gambling task [14]; in a similar study it prevented the rise in risk-taking that severe sleep loss otherwise caused [15]; and after almost two days without sleep, it restored alertness but not decision quality [16]. A review of caffeine's effects on the nervous system concluded that, overall, the evidence does not show that caffeine increases risk-taking or aggression [17]. A few experiments do find small shifts. In children and adolescents, 2 mg/kg increased risk-taking on a balloon-inflation task in moderate consumers but reduced it in heavy consumers [18]; an energy drink slightly increased risk-taking on the same task [19]; and 200 mg, though not 125 mg, made shoppers more easily swayed by a decoy product [20]. Effects also depend on habitual intake and physical state [21], and part of the apparent benefit of caffeine in regular drinkers is relief from overnight withdrawal [22].

Most of these experiments used tasks built to measure risk or inhibition, or tested people under extreme sleep loss. Less is known about whether a normal cup or two of coffee changes the way rested young adults respond in ordinary cognitive tasks, and few studies have asked whether feeling on edge actually goes with acting rashly. We previously studied a group of young adults who each received placebo, 70 mg and 140 mg of caffeine, and reported how dose and time after the drink affected performance, physiology and mood [23]. Cognitive benefits peaked near 1.2 mg/kg, and self-rated jitteriness after the tasks was higher at 140 mg. That analysis asked how well participants performed. The present study uses the same recordings to ask how they reached their answers.

We tested four signs of rashness: more errors, more guessing, a more trigger-happy response bias, and less time spent checking answers. A study of this size cannot prove that an effect is absent simply by failing to find one, so each comparison was paired with an equivalence test, which asks whether an

effect large enough to matter can be ruled out [24]. We also asked whether feeling jittery or anxious, or being physiologically aroused, went with rash responding within a session.

Methods

Study Design and Participants

This is a secondary analysis of a single-blind, counterbalanced crossover study whose protocol and main dose–time results have been reported [23]. The measures analysed here were not examined in that report. The study followed the Declaration of Helsinki and was approved by the Universiti Teknologi Malaysia Research Ethics Committee (UTMREC-2025-113). All participants gave written informed consent.

Eight university students aged 20 to 24 years took part (Table 1). One woman enrolled twice, a week apart, with a new intake assessment and a complete set of sessions each time. As in the earlier report, her two enrolments are treated as separate units, giving nine enrolments [23]. Body mass ranged from 47 to 113 kg. Habitual caffeine use was none for one enrolment, low for five and moderate for two, and was not reported for one. Participants were asked not to consume caffeine before each session. Symptoms of depression, anxiety and stress at intake were recorded with the DASS-21 [25].

Table 1

Participants and sessions. Doses are in mg per kg of body mass measured on the day. A dagger (†) marks a session without a physiological recording, which was left out of the main analysis but kept in the check using all 27 sessions. DASS-21 scores are the instrument's doubled sums (D, depression; A, anxiety; S, stress). E3 and E9 are the same woman, who enrolled twice.

Enrolment	Sex	Age (y)	Body mass (kg)	BMI (kg/m²)	Habitual caffeine	DASS-21 D / A / S	Placebo	70 mg	140 mg
E1	M	23	80	26.7	Low	6 / 4 / 2	0.025	0.875	1.750
E2	M	23	68	22.2	Low	2 / 10 / 8	0.029	1.029	2.059†
E3	F	24	47	20.9	Moderate	2 / 16 / 12	0.043	1.489	2.979
E4	M	21	53	20.0	Low	0 / 2 / 0	0.038	1.321	2.642
E5	M	21	113	39.1	Low	6 / 8 / 8	0.018	0.619	1.239
E6	M	21	57	19.7	Low	10 / 4 / 6	0.035	1.228	2.456†
E7	M	20	67	24.0	None	0 / 2 / 4	0.030	1.045	2.090
E8	M	20	63	21.8	Not reported	2 / 2 / 2	0.032	1.111	2.222
E9	F	24	48	21.3	Moderate	14 / 16 / 14	0.042	1.458	2.917

Caffeine and Session Procedure

Each participant attended three sessions on separate days. In each, they drank two instant-coffee sachets dissolved in warm water under observation: two decaffeinated sachets (about 2 mg; placebo), one decaffeinated and one regular sachet (70 mg), or two regular sachets (140 mg). The order was counterbalanced, and all six possible orders occurred. Participants were not told which drink they received; the investigator knew. Relative to body mass measured on the day, the doses were 0.018 to 0.043 mg/kg (placebo), 0.619 to 1.489 mg/kg (70 mg) and 1.239 to 2.979 mg/kg (140 mg). A test block was completed before the drink and again starting at a median of 5, 30 and 55 min after it.

Tasks and Ratings

Each block began with mood ratings on a 0 to 100 visual analogue scale (VAS), followed by three computer tasks run on Testable [26], with new stimuli in every block.

In the flanker task [27], participants pressed a key to show which way a central arrow or letter was pointing while ignoring the symbols beside it, which pointed the same way, the opposite way or neither way. There were 48 trials per block. In the visual short-term memory (VSTM) task [28], three to five coloured objects appeared briefly (100 ms), disappeared for 0.9 s and reappeared; participants answered whether anything had changed. Half of the 24 trials contained a change. In serial subtraction by sevens (SS7), participants typed answers counting down by seven from a starting number for 60 s, pressing Enter after each answer.

What Was Measured

Errors. The proportion of wrong answers on the flanker task and on SS7. For SS7, the last answer of each block was often cut off by the 60 s timer, sometimes after a digit or two had been typed; these unfinished answers are not decisions and were left out.

Guessing. A response faster than 300 ms is too quick for the flanker display to have been properly seen, so such responses were counted as guesses. We also looked at the error rate in each person's fastest fifth of responses, because a person who guesses makes most of their mistakes when responding fastest, and at the response time of each block's fastest 10% of responses, which would shorten if people began answering before looking properly.

Response bias. When people are unsure, they tend to lean one way. In the VSTM task, a participant unsure whether anything changed can play safe and answer "no change" or jump and answer "changed". Signal detection theory separates how well a person can actually see a change (sensitivity, d') from which way they lean when unsure (response bias, c) [29]. A positive c means a cautious habit of answering "no change", and a negative c a trigger-happy habit of answering "changed". If caffeine made people rash, c should fall.

Checking answers. Testable recorded the time of every keystroke in SS7. We measured the pause between typing the last digit of an answer and pressing Enter, the moment at which a participant can look over an answer before committing to it. Someone who rushes would shorten this pause. The time taken to type the first digit was also recorded. Because every digit typed was logged, including digits later deleted, an answer with more keystrokes than final digits had been corrected before it was submitted; the share of self-corrected answers is a second, direct measure of checking.

What the errors looked like. To see whether errors were guesses or hard calls, each error was placed in its own block's ranking of response times, from fastest (0) to slowest (1); guesses would cluster among the fastest. Each person's median error time was compared with their median correct time, using a Wilcoxon signed-rank test across enrolments, and on the flanker task the three trials before each error were compared with the three before each correct answer, to see whether people sped up into their mistakes. VSTM errors were split into misses (answering "no change" when something had changed) and false alarms (answering "changed" when nothing had). Wrong SS7 answers were sorted by the digits

typed into entry slips (a missing or extra digit, two digits swapped, or a wrong first digit) and calculation slips (an answer off by exactly 10 or 100, or wrong in its last digits).

Speed and mood. Average response time on correct trials was kept to show whether caffeine changed speed. Mood was taken from the ratings made just before each block's tasks (jitteriness, anxiety, stress, alertness, sleepiness, mental fatigue and motivation), because that is how participants felt while making that block's decisions. Jitteriness rated after the tasks, the measure used in the earlier report, was added for comparison.

Physiological Recording

Skin conductance and the pulse were recorded throughout each session with a Shimmer3 GSR+ unit [30] and processed as described previously [23]. For each block we calculated the skin conductance level, the rate of spontaneous skin conductance responses and the mean heart rate. Blocks with less than half of the signal usable were excluded.

Sessions Included

The main analysis used the same 25 sessions as the earlier report. Two sessions had no physiological recording and four post-dose blocks lost their recording; these were excluded there. Because all of these sessions and blocks have complete task and mood data, every analysis was repeated on all 27 sessions as a check. One pre-dose flanker block (in a placebo session) was not recorded, so that session was left out of the flanker comparisons.

Statistical Analysis

Why scores were not simply subtracted. A common way to analyse a before-and-after design is to subtract each session's pre-dose score from its post-dose score. This can mislead when the groups happen to start at different levels, and here they did. By chance, the placebo sessions started with far more flanker errors than the others: 16 of the 62 flanker errors in the whole study were made in placebo pre-dose blocks, against 2 in 70 mg pre-dose blocks. An unusually poor start tends to be followed by a more ordinary score anyway, a pattern known as regression to the mean. Subtraction would therefore have made placebo look as though it improved and caffeine look as though it caused more errors, when after the drink the caffeine sessions actually made fewer errors. Instead, we compared post-dose scores directly while adjusting for each session's pre-dose score (analysis of covariance), the recommended approach for this situation [31].

The comparison. Each post-dose block was compared across drinks using a linear mixed model, a form of regression that gives each enrolment its own starting level, so that each person is compared with themselves. The model included the drink, the session's pre-dose score, the time of the block and the visit number (to allow for practice). Because the data come from only nine enrolments, confidence

intervals and p-values were made appropriately wider by using a t distribution with eight degrees of freedom [32].

A check that needs no assumptions. Each comparison was repeated as a permutation test. If caffeine had no effect, the labels "placebo", "70 mg" and "140 mg" would be interchangeable within each person. The labels were shuffled at random within each participant who completed all three drinks, 20,000 times, and the permutation p-value is the proportion of shuffles that produced a difference at least as large as the real one.

Ruling out effects that matter. A non-significant result does not show that an effect is absent; the study may simply have been too small to detect it. Equivalence testing deals with this [24]. The smallest effect that would matter is set first, as an equivalence bound, and the data are then checked to see whether an effect at least that large can be ruled out. No accepted values exist for these measures, so each bound was set at half the standard deviation of that measure across all pre-dose blocks, a conventional medium-sized effect. For measures where rashness has a direction, such as more errors or a lower c, we asked the one-sided question: can a rash effect as large as the bound be ruled out? This is the case when the upper end of the 90% confidence interval lies inside the bound. The bounds were set after the data had been collected, so these tests are presented as descriptive.

Feeling and responding. To see whether feeling on edge went with responding rashly, each block's ratings and physiology were correlated with that block's task performance. Each session's average and the average of each time point were removed first, so that a correlation reflects only whether a block that was unusually tense was also unusually error-prone, rather than differences between people or the general drift across a session [33]. Five state measures (heart rate, skin conductance level, skin conductance response rate, jitteriness and anxiety) were paired with five performance measures, giving 25 correlations. With that many tests some will reach $p < 0.05$ by chance, so the p-values were adjusted for the number of tests [34].

Results

Data

The main analysis included 4,559 flanker trials, 2,304 VSTM trials and 94 SS7 blocks. Participants were accurate on the flanker task, with 98.6% of responses correct. Table 2 shows each measure before and after the drink.

Table 2

Each measure before the drink (the pre-dose block) and after it (the three post-dose blocks together), main analysis. Values are mean (standard deviation) across blocks. The first row gives the number of blocks; one placebo pre-dose flanker block was not recorded, so flanker values before placebo come from 8 blocks, and two post-dose SS7 blocks were not recorded, so SS7 values after placebo and 70 mg come from 25 and 24 blocks.

Measure	Placebo before	Placebo after	70 mg before	70 mg after	140 mg before	140 mg after
Blocks (n)	9	26	9	25	7	20
Flanker error rate (%)	4.2 (2.5)	1.6 (1.9)	0.5 (0.9)	0.8 (1.3)	2.1 (2.4)	0.7 (1.6)
Flanker response time (ms)	548 (95)	566 (125)	666 (211)	592 (139)	663 (145)	633 (126)
Flanker fastest 10% of responses (ms)	379 (54)	399 (45)	414 (52)	410 (52)	409 (51)	407 (44)
VSTM response bias c	0.30 (0.47)	0.42 (0.32)	0.41 (0.35)	0.40 (0.36)	0.42 (0.10)	0.45 (0.29)
VSTM sensitivity d'	1.99 (0.46)	1.85 (0.46)	1.87 (0.83)	2.12 (0.60)	1.86 (0.70)	2.02 (0.64)
SS7 error rate (%)	2.5 (3.9)	5.8 (8.0)	10.7 (11.4)	4.3 (7.0)	1.0 (2.7)	4.7 (8.0)
SS7 pause before Enter (ms)	553 (294)	528 (228)	563 (247)	533 (237)	669 (534)	556 (228)
SS7 answers self-corrected (%)	6.0 (6.1)	7.7 (7.1)	15.5 (18.5)	7.4 (6.5)	15.6 (12.5)	8.3 (9.0)
Jitteriness before block (0–100)	14.8 (25.8)	10.7 (16.3)	12.6 (12.2)	12.4 (21.0)	8.4 (14.8)	9.3 (18.0)
Jitteriness after block (0–100)	18.2 (28.1)	13.4 (20.9)	15.3 (22.8)	11.8 (19.5)	6.0 (8.7)	9.4 (14.6)
Anxiety before block (0–100)	12.0 (18.8)	9.1 (14.0)	7.8 (11.1)	6.7 (10.0)	7.6 (14.5)	8.6 (17.7)
Sleepiness before block (0–100)	28.9 (31.1)	22.2 (26.8)	29.3 (25.4)	19.0 (22.6)	27.3 (27.3)	8.7 (12.3)

Errors and Guessing

Errors after the drink were rare in every arm: 1.6% of flanker responses with placebo, 0.8% with 70 mg and 0.7% with 140 mg (Fig. 1a). Before the drink the same arms had made 4.2%, 0.5% and 2.1% errors. After adjusting for each session's pre-dose errors, 140 mg produced 1.15 percentage points fewer errors than placebo and 70 mg 0.77 fewer, but neither difference was statistically significant (p = 0.082 and 0.31; Table 3). In the check that included all 27 sessions, the 140 mg reduction was 1.14 percentage

points and was significant (95% confidence interval 0.06 to 2.22 points fewer; p = 0.041). SS7 errors did not differ between drinks (Fig. 1c; Table 3).

Table 3

Difference from placebo after the drink, adjusted for each session's pre-dose score (main analysis). Estimate: caffeine minus placebo, with its 95% confidence interval. p: from the mixed model. Permutation p: from shuffling the drink labels within each person. Bound: the equivalence bound, half the standard deviation across pre-dose blocks. Rash effect ruled out (p): one-sided test that a change in the rash direction (more errors, a lower c, faster fastest responses, a shorter pause before Enter, fewer self-corrections, more jitteriness, anxiety or stress) is at least as large as the bound; p < 0.05 means such an effect is ruled out. A dash marks measures with no rash direction.

Measure	Bound	Drink	Estimate (95% CI)	p	Permutation p	Rash effect ruled out (p)
Flanker error rate (percentage points)	1.25	70 mg	−0.77 (− 2.42 to 0.88)	0.31	0.84	0.011
		140 mg	−1.15 (− 2.49 to 0.18)	0.082	0.087	0.002
VSTM response bias c	0.17	70 mg	−0.01 (− 0.19 to 0.17)	0.88	0.93	0.035
		140 mg	0.07 (− 0.13 to 0.27)	0.45	0.41	0.011
SS7 error rate (percentage points)	4.15	70 mg	−2.5 (− 8.2 to 3.3)	0.35	0.80	0.014
		140 mg	−1.4 (− 6.6 to 3.9)	0.57	0.54	0.021
Flanker response time (ms)	82	70 mg	−2 (− 64 to 60)	0.95	0.90	–
		140 mg	26 (− 41 to 93)	0.40	0.40	–
Flanker fastest 10% of responses (ms)	26	70 mg	3 (− 12 to 17)	0.69	0.76	< 0.001
		140 mg	−7 (− 22 to 8)	0.30	0.43	0.010
VSTM sensitivity d'	0.33	70 mg	0.30 (− 0.02 to 0.61)	0.061	0.034	–
		140 mg	0.21 (− 0.13 to 0.55)	0.19	0.21	–
SS7 pause before Enter (ms)	175	70 mg	−3 (− 95 to 88)	0.93	0.96	0.001
		140 mg	−25 (− 135 to 84)	0.60	0.63	0.007
SS7 answers self-corrected (percentage points)	6.85	70 mg	−0.02 (− 4.8 to 4.8)	0.99	0.54	0.006

Measure	Bound	Drink	Estimate (95% CI)	p	Permutation p	Rash effect ruled out (p)
		140 mg	−0.2 (− 5.3 to 5.0)	0.93	0.96	0.009
Jitteriness before block	9.1	70 mg	3.7 (− 6.4 to 13.8)	0.42	0.59	0.13
		140 mg	2.8 (− 8.0 to 13.7)	0.57	0.67	0.11
Jitteriness after block	11.0	70 mg	−2.0 (− 8.4 to 4.4)	0.49	0.19	< 0.001
		140 mg	6.1 (− 1.7 to 13.8)	0.11	0.87	0.092
Anxiety before block	7.3	70 mg	1.2 (− 4.2 to 6.7)	0.62	0.94	0.016
		140 mg	3.4 (− 2.3 to 9.1)	0.20	0.34	0.076
Stress before block	8.4	70 mg	1.6 (− 4.0 to 7.2)	0.53	0.89	0.011
		140 mg	4.2 (− 1.8 to 10.3)	0.15	0.42	0.073
Sleepiness before block	13.5	70 mg	−3.0 (− 14.2 to 8.1)	0.55	0.48	–
		140 mg	−8.2 (− 20.4 to 4.0)	0.16	0.099	–

Guessing was almost absent. Only 11 of the 4,559 flanker responses were faster than 300 ms, six of them in placebo pre-dose blocks, and the fastest response was 199 ms. If caffeine had encouraged guessing, errors would have climbed among the fastest responses after the drink. The opposite was seen: in each person's fastest fifth of responses after the drink, errors were 2.2% with placebo (6 of 273), 1.4% with 70 mg (3 of 213) and 0.5% with 140 mg (1 of 204) (Fig. 2). Nor did the fastest responses get quicker: the fastest 10% of each block's responses took about as long after caffeine as after placebo (3 ms longer after 70 mg and 7 ms shorter after 140 mg).

Response Bias

Participants were cautious change detectors. They answered "no change" on 61.0% of VSTM trials, although only 50% of trials had no change, and their average c was positive in every arm before and after the drink (Table 2). Caffeine did not shift this habit towards answering "changed": c changed by − 0.01

after 70 mg and + 0.07 after 140 mg, both close to zero (Fig. 1b). Sensitivity improved slightly instead, with d' rising by 0.30 after 70 mg (p = 0.061; permutation p = 0.034) and by 0.21 after 140 mg (p = 0.19).

Speed and Checking

Caffeine did not change average flanker response time (− 2 ms after 70 mg and + 26 ms after 140 mg), which stayed within its equivalence bound of ± 82 ms. The pause before pressing Enter in SS7 did not shorten (− 3 ms after 70 mg and − 25 ms after 140 mg) and stayed within its bound of ± 175 ms. Self-corrections were equally common after every drink, at 7.7% of answers with placebo, 7.4% with 70 mg and 8.3% with 140 mg (Table 2), and a fall as large as the bound was ruled out (Table 3). The time to type the first digit varied too much between blocks to draw a conclusion.

Can a Rash Effect Be Ruled Out?

For all three main measures, a rash effect of the size of the equivalence bound was ruled out at both doses (Fig. 3; Table 3): an increase in flanker errors (p = 0.011 for 70 mg and p = 0.002 for 140 mg), a shift towards trigger-happy answering on the VSTM task (p = 0.035 and 0.011) and an increase in SS7 errors (p = 0.014 and 0.021). Rash changes in the speed of the fastest responses, in the pause before Enter and in self-corrections were also ruled out. The same conclusions held when all 27 sessions were included.

What the Errors Looked Like

The errors themselves did not look like guesses (Online Resource 1). On the flanker task, errors were twice as common when the side arrows pointed the opposite way (2.1% of those trials) as when they pointed the same way (1.0%), and almost absent when they pointed neither way (0.4%), so they followed what was on the screen. They were not unusually fast: within its block, the typical error sat in the middle of the speed ranking, and 27% of errors fell among the block's fastest fifth of responses, close to the 20% expected by chance. Participants did not speed up in the trials leading to an error (p = 0.13), and blocks in which a person responded faster did not contain more errors (r = − 0.14, p = 0.21). On the memory task, errors were slower than correct answers, by a median of 103 ms per person (p = 0.008), which is the pattern of hesitation rather than guessing, and most of them were misses, the cautious answer "no change" given when something had changed. After the drink, false alarms, the trigger-happy error, numbered 22 with placebo, 17 with 70 mg and 12 with 140 mg. In serial subtraction, about half of the 58 wrong answers were entry slips, most often a dropped zero (typing 37 for 307), and the rest were almost all calculation slips. Neither kind became more common after caffeine, and wrong answers were typed and checked as slowly as correct ones.

How Participants Felt

Jitteriness rated before each block did not differ from placebo (+ 3.7 points after 70 mg and + 2.8 after 140 mg on the 0 to 100 scale), and neither did anxiety or stress (Table 3). After 140 mg, however, increases of the size of the bound could not be ruled out for any of these three ratings, so small rises

remain possible. Jitteriness rated after the tasks, the measure used in the earlier report, was 6.1 points higher after 140 mg than after placebo once the pre-dose rating was taken into account, a difference that was not significant (p = 0.11). Sleepiness was lower after 140 mg (8.2 points; p = 0.16), and this reduction was significant when all 27 sessions were included (13.4 points; p = 0.026). Alertness, mental fatigue and motivation did not differ from placebo.

Feeling, Arousal and Responding

Within a session, how jittery or anxious participants felt before a block was unrelated to how they performed in it (Table 4). The correlation coefficient r measures how closely two measures move together, from − 1 to + 1, with 0 meaning no relationship; for jitteriness and anxiety, every r lay between − 0.17 and 0.15. Skin conductance level and the rate of skin conductance responses were also unrelated to performance (r between − 0.12 and 0.13). Blocks with a higher heart rate than the session's own average had fewer flanker errors (r = − 0.31, p = 0.011) and slower flanker responses (r = 0.26, p = 0.033). Once the p-values were adjusted for the 25 correlations tested, however, neither association remained reliable (adjusted p, or q, = 0.28 and 0.42, against the usual threshold of 0.05).

Table 4

Do feeling and arousal go with how people respond? Each cell gives the correlation r and its p-value in brackets, calculated within sessions after removing each session's average and the average of each time point (91 to 96 blocks from 25 sessions). r runs from − 1 to + 1; 0 means no relationship, and a negative value means that when one measure was higher than usual the other was lower. After adjusting the p-values for the 25 correlations tested (q), none was reliable: the smallest q was 0.28, against the usual threshold of 0.05.

State	Flanker error rate	Flanker response time	VSTM response bias c	VSTM sensitivity d'	SS7 pause before Enter
Heart rate	−0.31 (0.011)	0.26 (0.033)	0.16 (0.20)	−0.04 (0.77)	−0.05 (0.71)
Skin conductance level	−0.12 (0.32)	0.09 (0.44)	−0.02 (0.89)	0.13 (0.29)	0.02 (0.89)
Skin conductance response rate	−0.10 (0.40)	0.02 (0.86)	0.07 (0.57)	0.11 (0.38)	−0.05 (0.69)
Jitteriness before block	0.00 (0.99)	−0.05 (0.67)	−0.09 (0.45)	−0.03 (0.78)	−0.17 (0.19)
Anxiety before block	0.15 (0.22)	−0.07 (0.59)	−0.06 (0.64)	0.14 (0.25)	−0.14 (0.27)

Discussion

In this pilot study, everyday doses of caffeine did not make decisions more hasty or error-prone. Errors did not rise on any task, and flanker errors were, if anything, lowest after 140 mg. Participants did not guess more, did not become more willing to report changes they had not seen and did not rush their arithmetic answers. For each of the three main measures, an increase in rashness of half a standard

deviation could be ruled out at both doses. The errors themselves looked like hard calls rather than guesses: they followed what was on the screen, were no faster than correct answers and, on the memory task, were slower. How participants felt and how they performed also proved to be separate: feeling more jittery or anxious before a block did not go with responding more rashly in it.

These results agree with most experiments in which caffeine is given directly. With a dose similar to our highest, Tieges and colleagues found no weakening of the ability to hold back responses on flanker and stop-signal tasks, only a general speeding of responses [13]. In studies of severe sleep loss, caffeine did not worsen decisions on a gambling task [14, 16] and prevented the increase in risk-taking that sleep loss otherwise produced [15]. A review of caffeine's effects on behaviour reached the same overall conclusion [17]. Studies that reported more risk-taking after caffeine involved a balloon-inflation task in children who were moderate consumers [18], an energy drink containing other ingredients [19] or consumer choices after 200 mg [20], and the shifts were small. In the language of the speed–accuracy trade-off [9], a rash decision-maker gives up accuracy for speed, whereas a sharper one gains accuracy, speed or both. Our participants did not get faster on the flanker task but tended to make fewer errors, and on the memory task they detected changes better without changing their bias. Both point to clearer processing rather than a lower threshold for answering, in line with reports that caffeine improves the control of attention [35] and may reduce errors in shift workers [36].

How, then, do surveys come to link caffeine with impulsivity [10–12]? The most likely answer is that they describe people rather than doses. Sensation seekers and impulsive people may simply consume more caffeine, just as the heavier caffeine users in the gambling study also smoked more [11], without each cup making them more impulsive. Grant and Chamberlain themselves noted that heavy caffeine use was not related to motor inhibition in their sample [11], which is the form of rashness our tasks measured. The two kinds of evidence are therefore not in conflict: a habit of high caffeine intake can mark an impulsive personality, while a single everyday dose does not make a rested person rash.

The separation between feeling on edge and acting rashly matters for how caffeine's mental effects are understood. Jitteriness is the unpleasant effect that limits how much caffeine people take [5], and our earlier report found it higher after the tasks at 140 mg [23]. When the pre-dose rating was taken into account, that difference was about half as large and was no longer statistically reliable, because the 140 mg sessions happened to start with the lowest ratings. Whatever its true size, the feeling did not carry over into behaviour: blocks in which participants felt more jittery were no more error-prone, biased or hurried than other blocks. Caffeine's effects on how people feel and how they perform have been separated before, notably in work showing that much of its benefit in regular drinkers is relief from withdrawal [22]. Our results add that, up to about 3 mg/kg, the feeling of being on edge does not appear to reach the quality of decisions.

The physiological recordings offered one lead. Within a session, blocks with a higher heart rate went with fewer errors and slower responses, a pattern of care rather than haste. The link did not survive adjustment for the number of tests, and heart rate also falls as participants settle into a session, so it is

a question for future work rather than a finding. Skin conductance, whose changes tracked stress and anxiety ratings in the earlier report [23], was unrelated to performance. Errors themselves trigger a small autonomic response [37], and measuring that response after each individual error, with and without caffeine, would test the link more directly than block averages; it would require the task and sensor clocks to be aligned to within a fraction of a second, which this study did not achieve.

The study also illustrates a practical point for small before-and-after studies. When the same data were analysed by subtracting pre-dose from post-dose scores, caffeine appeared to increase flanker errors by 2.9 percentage points after 70 mg ($p < 0.001$) and very fast responses by 1.3 points at both doses ($p \leq 0.018$). Both effects were artefacts of the placebo sessions' poor start, and they reversed once the pre-dose score was used as an adjustment rather than subtracted [31]. The same adjustment roughly halved the jitteriness effect reported earlier. With only seven to nine sessions per drink, this protection matters.

The study has limitations. The sample is small, with nine enrolments of eight people, and the equivalence bounds were chosen after the data were collected. The confidence intervals leave room for modest effects in either direction, so the conclusion is that large rash effects are unlikely, not that caffeine has no effect at all. Accuracy on the flanker task was close to perfect, so the task could only reveal rash errors if they were common; tasks built to provoke impulsive responses, such as stop-signal or gambling tasks, would test judgement more directly. The doses were everyday ones, at most 140 mg or about 3 mg/kg, within the range considered safe for a single dose in adults [38], and rash decision-making may appear at higher doses, under sleep loss or in people prone to anxiety. Participants were young, mostly men and mostly light consumers, and the investigator knew which drink was given. Finally, this is a secondary analysis of a cohort whose main results have been reported [23], although the measures and the question are new.

Conclusion

Everyday doses of caffeine, up to 140 mg or about 3 mg/kg, did not make young adults' decisions more hasty or error-prone. After the drink, flanker error rates were 1.6% with placebo, 0.8% with 70 mg and 0.7% with 140 mg, and at both doses an increase in rashness of half a standard deviation was ruled out for flanker errors, arithmetic errors and trigger-happy answering on the memory task ($p \leq 0.035$). The errors that did occur were hard calls rather than guesses, and answer checking did not decline. Feeling on edge and acting rashly appeared to be separate. A larger study using tasks designed to provoke impulsive responses is needed to confirm that caffeine leaves judgement intact.

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Universiti Teknologi Malaysia Research Ethics Committee (UTMREC-2025-113). All participants gave written informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the Malaysia Ministry of Higher Education and Universiti Teknologi Malaysia under UTM Fundamental Research (PY/2023/00971; Cost Centre Q.J130000.3823.23H11).

Author Contribution

R.T.Y.S.: conceptualisation, methodology, investigation, data curation, formal analysis, visualisation, writing – original draft. R.S.: conceptualisation, supervision, funding acquisition, writing – review and editing. E.S.L.M.: supervision, writing – review and editing. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the participants for their time.

Data Availability

De-identified data and analysis code are available from the corresponding author on reasonable request, subject to approval by the Universiti Teknologi Malaysia Research Ethics Committee.

References

1. Galderisi S, Heinz A, Kastrup M, Beezhold J, Sartorius N. Toward a new definition of mental health. World Psychiatry. 2015;14(2):231–3. https://doi.org/10.1002/wps.20231.
2. Smith A. Effects of caffeine on human behavior. Food Chem Toxicol. 2002;40(9):1243–55. https://doi.org/10.1016/s0278-6915(02)00096-0.
3. McLellan TM, Caldwell JA, Lieberman HR. A review of caffeine's effects on cognitive, physical and occupational performance. Neurosci Biobehav Rev. 2016;71:294–312. https://doi.org/10.1016/j.neubiorev.2016.09.001.

4. Lara DR. Caffeine, mental health, and psychiatric disorders. J Alzheimers Dis. 2010;20(Suppl 1):S239–48. https://doi.org/10.3233/jad-2010-1378.
5. Childs E, de Wit H. Subjective, behavioral, and physiological effects of acute caffeine in light, nondependent caffeine users. Psychopharmacology. 2006;185(4):514–23. https://doi.org/10.1007/s00213-006-0341-3.
6. Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit-formation. J Comp Neurol Psychol. 1908;18(5):459–82. https://doi.org/10.1002/cne.920180503.
7. Revelle W, Humphreys MS, Simon L, Gilliland K. The interactive effect of personality, time of day, and caffeine: a test of the arousal model. J Exp Psychol Gen. 1980;109(1):1–31. https://doi.org/10.1037/0096-3445.109.1.1.
8. Grant JE, Chamberlain SR. Impulsive action and impulsive choice across substance and behavioral addictions: cause or consequence? Addict Behav. 2014;39(11):1632–9. https://doi.org/10.1016/j.addbeh.2014.04.022.
9. Heitz RP. The speed-accuracy tradeoff: history, physiology, methodology, and behavior. Front Neurosci. 2014;8:150. https://doi.org/10.3389/fnins.2014.00150.
10. Jones HA, Lejuez CW. Personality correlates of caffeine dependence: the role of sensation seeking, impulsivity, and risk taking. Exp Clin Psychopharmacol. 2005;13(3):259–66. https://doi.org/10.1037/1064-1297.13.3.259.
11. Grant JE, Chamberlain SR. Caffeine's influence on gambling behavior and other types of impulsivity. Addict Behav. 2018;76. https://doi.org/10.1016/j.addbeh.2017.08.007. :156 – 60.
12. Shima C, Lee R, Coccaro EF. Associations of agression and use of caffeine, alcohol and nicotine in healthy and aggressive individuals. J Psychiatr Res. 2022;146:21–7. https://doi.org/10.1016/j.jpsychires.2021.10.015.
13. Tieges Z, Snel J, Kok A, Ridderinkhof KR. Caffeine does not modulate inhibitory control. Brain Cogn. 2009;69(2):316–27. https://doi.org/10.1016/j.bandc.2008.08.001.
14. Killgore WDS, Lipizzi EL, Kamimori GH, Balkin TJ. Caffeine effects on risky decision making after 75 hours of sleep deprivation. Aviat Space Environ Med. 2007;78(10):957–62. https://doi.org/10.3357/asem.2106.2007.
15. Killgore WDS, Kamimori GH, Balkin TJ. Caffeine protects against increased risk-taking propensity during severe sleep deprivation. J Sleep Res. 2011;20(3):395–403. https://doi.org/10.1111/j.1365-2869.2010.00893.x.
16. Killgore WDS, Grugle NL, Balkin TJ. Gambling when sleep deprived: don't bet on stimulants. Chronobiol Int. 2012;29(1):43–54. https://doi.org/10.3109/07420528.2011.635230.
17. Turnbull D, Rodricks JV, Mariano GF. Neurobehavioral hazard identification and characterization for caffeine. Regul Toxicol Pharmacol. 2016;74:81–92. https://doi.org/10.1016/j.yrtph.2015.12.002.
18. Temple JL, Ziegler AM, Graczyk AM, Crandall A. Effects of acute and chronic caffeine on risk-taking behavior in children and adolescents. J Psychopharmacol. 2017;31(5):561–8. https://doi.org/10.1177/0269881117691568.

19. Peacock A, Bruno R, Martin FH, Carr A. The impact of alcohol and energy drink consumption on intoxication and risk-taking behavior. Alcohol Clin Exp Res. 2013;37(7):1234–42. https://doi.org/10.1111/acer.12086.
20. Canty M, Lang FJ, Adler SJ, Lichters M, Sarstedt M. Caffeine's complex influence on the attraction effect: a mixed bag of outcomes. Mark Lett. 2024;35(3):451–76. https://doi.org/10.1007/s11002-023-09710-6.
21. Huertas F, Blasco E, Moratal C, Lupiañez J. Caffeine intake modulates the functioning of the attentional networks depending on consumption habits and acute exercise demands. Sci Rep. 2019;9(1):10043. https://doi.org/10.1038/s41598-019-46524-x.
22. Rogers PJ, Heatherley SV, Mullings EL, Smith JE. Faster but not smarter: effects of caffeine and caffeine withdrawal on alertness and performance. Psychopharmacology. 2013;226(2):229–40. https://doi.org/10.1007/s00213-012-2889-4.
23. Teoh RYS, Sudirman R, Su ELM. Caffeine-intake variables based on physiological indicators for cognitive function using response surface methodology: a body-mass-scaled dose-time optimisation. In: Proceedings of the 15th International Conference on Bioinformatics and Biomedical Science (ICBBS 2026); 2026 Oct 16–18; Seoul, Republic of Korea. New York: ACM; 2026. In press.
24. Lakens D, Scheel AM, Isager PM. Equivalence testing for psychological research: a tutorial. Adv Methods Pract Psychol Sci. 2018;1(2):259–69. https://doi.org/10.1177/2515245918770963.
25. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol. 2005;44(2):227–39. https://doi.org/10.1348/014466505x29657.
26. Rezlescu C, Danaila I, Miron A, Amariei C. More time for science: using Testable to create and share behavioral experiments faster, recruit better participants, and engage students in hands-on research. Prog Brain Res. 2020;253:243–62. https://doi.org/10.1016/bs.pbr.2020.06.005.
27. Eriksen BA, Eriksen CW. Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept Psychophys. 1974;16(1):143–9. https://doi.org/10.3758/bf03203267.
28. Luck SJ, Vogel EK. The capacity of visual working memory for features and conjunctions. Nature. 1997;390(6657):279–81. https://doi.org/10.1038/36846.
29. Stanislaw H, Todorov N. Calculation of signal detection theory measures. Behav Res Methods Instrum Comput. 1999;31(1):137–49. https://doi.org/10.3758/bf03207704.
30. Burns A, Greene BR, McGrath MJ, O'Shea TJ, Kuris B, Ayer SM, et al. SHIMMER™ – a wireless sensor platform for noninvasive biomedical research. IEEE Sens J. 2010;10(9):1527–34. https://doi.org/10.1109/jsen.2010.2045498.
31. Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. BMJ. 2001;323(7321):1123–4. https://doi.org/10.1136/bmj.323.7321.1123.
32. Cameron AC, Miller DL. A practitioner's guide to cluster-robust inference. J Hum Resour. 2015;50(2):317–72. https://doi.org/10.3368/jhr.50.2.317.

33. Bakdash JZ, Marusich LR. Repeated measures correlation. Front Psychol. 2017;8:456. https://doi.org/10.3389/fpsyg.2017.00456.
34. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol. 1995;57(1):289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
35. Brunyé TT, Mahoney CR, Lieberman HR, Giles GE, Taylor HA. Acute caffeine consumption enhances the executive control of visual attention in habitual consumers. Brain Cogn. 2010;74(3):186–92. https://doi.org/10.1016/j.bandc.2010.07.006.
36. Ker K, Edwards PJ, Felix LM, Blackhall K, Roberts I. Caffeine for the prevention of injuries and errors in shift workers. Cochrane Database Syst Rev. 2010;5CD008508. https://doi.org/10.1002/14651858.cd008508.
37. Hajcak G, McDonald N, Simons RF. To err is autonomic: error-related brain potentials, ANS activity, and post-error compensatory behavior. Psychophysiology. 2003;40(6):895–903. https://doi.org/10.1111/1469-8986.00107.
38. EFSA Panel on Dietetic Products, Nutrition and Allergies (NDA). Scientific opinion on the safety of caffeine. EFSA J. 2015;13(5):4102. https://doi.org/10.2903/j.efsa.2015.4102.

Figures

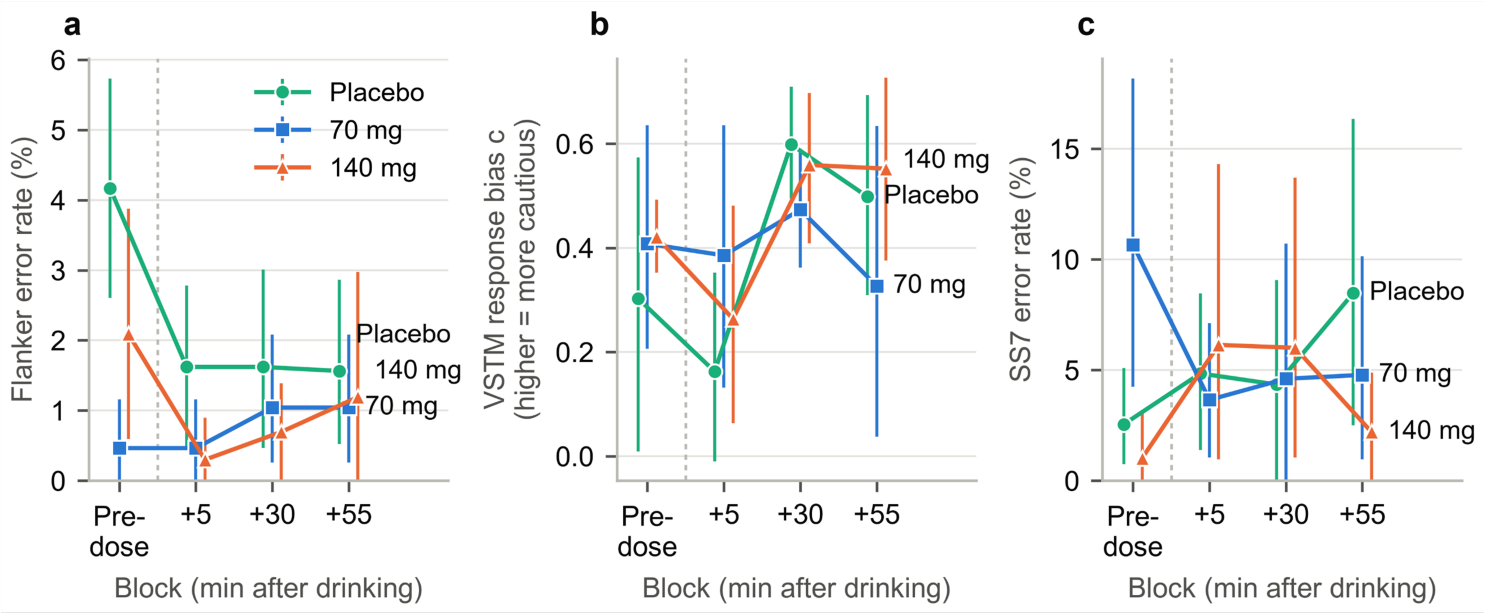


Figure 1

The three main measures across the session, by drink: (a) flanker error rate, (b) VSTM response bias c, where higher means a stronger habit of answering "no change", and (c) serial-subtraction error rate. Points are block averages with 95% confidence intervals; the dotted line separates the pre-dose block from the blocks 5, 30 and 55 min after the drink. The placebo sessions started with the most flanker

errors, which is why post-dose scores were adjusted for the pre-dose score rather than having it subtracted



Figure 2

Flanker errors by response speed, (a) in the pre-dose block and (b) in the three blocks after the drink. Each person's responses were split into five equal groups from fastest to slowest. Guessing would show up as more errors among the fastest responses after caffeine; instead, the fastest responses after 140 mg contained the fewest errors

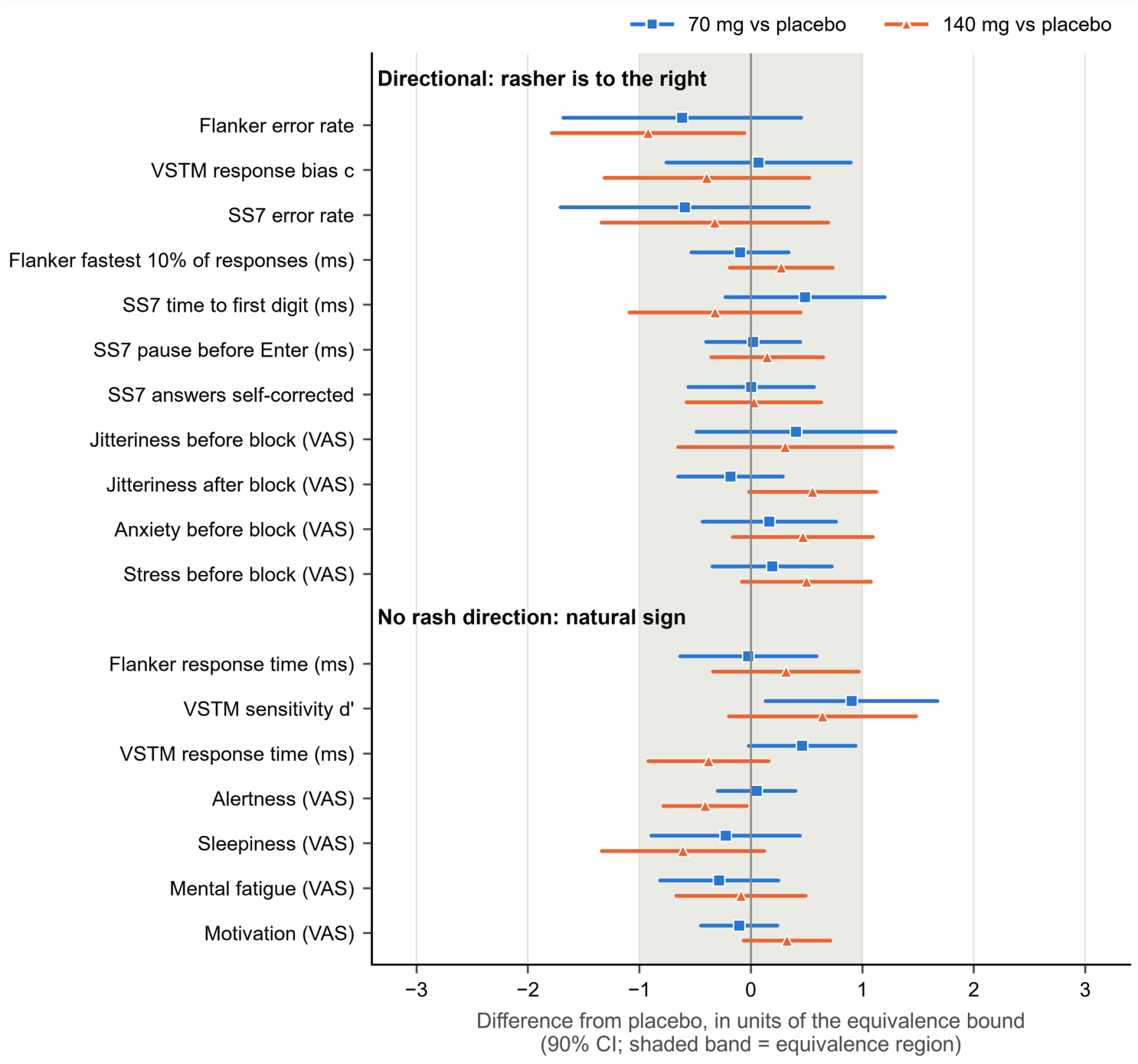


Figure 3

Every caffeine–placebo difference, expressed in units of its own equivalence bound, with 90% confidence intervals. The shaded band from −1 to +1 is the range of differences too small to matter. In the upper group, measures are lined up so that a rasher response lies to the right; a line that ends to the left of +1 means a rash effect of that size is ruled out. Measures in the lower group have no rash direction and keep their natural sign

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.docx