Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-024-02547-8

BRIEF REPORT

Context effects in cognitive effort evaluation

Sophie Desjardins[1] · Rui Tang[1] · Seffie Yip[1] · Mathieu Roy[1] · A. Ross Otto[1]

Accepted: 7 July 2024

Abstract
When given a choice, people will avoid cognitively effortful courses of action because the experience of effort is evaluated as aversive and costly. At the same time, a body of work spanning psychology, economics, and neuroscience suggests that goods, actions, and experiences are often evaluated in the context in which they are encountered, rather in absolute terms. To probe the extent to which the evaluation of cognitive effort is also context-dependent, we had participants learn associations between unique stimuli and subjective demand levels across low-demand and high-demand contexts. We probed demand preferences and subjective evaluation using a forced-choice paradigm as well by examining effort ratings, taken both on-line (during learning) and off-line (after choice). When choosing between two stimuli objectively identical in terms of demand, participants showed a clear preference for the stimulus learned in the low- versus high-demand context and rated this stimulus as more subjectively effortful than the low-demand context in on-line but not off-line ratings, suggesting an assimilation effect. Finally, we observed that the extent to which individual participants who exhibited stronger assimilation effects in off-line demand ratings were more likely to manifest an assimilation effect in demand preferences. Broadly, our findings suggest that effort evaluations occur in a context-dependent manner and are specifically assimilated to the broader context in which they occur.

Keywords Cognitive effort · Cognitive demand · Context effects · Decision-making

Introduction

Cognitive effort is aversive—when possible, people will often choose to avoid exerting effort. Underlying this point, individuals will choose to receive physical pain in order to avoid performing a demanding cognitive task (Vogel et al., 2020). Moreover, the experience of cognitive effort exertion is believed to carry disutility—that is, it feels aversive and costly (Chen et al., 2023; Devine et al., 2023). As epitomized by Hull's (1943) classic "law of less work", individuals, when given the choice, tend to favour less cognitively demanding over more demanding cognitive courses of action (Kool et al., 2010).

While the cognitive effort literature has examined individuals' evaluations of cognitive effort costs using, for example, self-report scales and preferences between effort levels (Devine & Otto, 2022; Strobel et al., 2020; Vermeylen et al., 2022; Westbrook et al., 2013), this work typically takes these subjective valuations and preferences as absolute measures, independent of the context in which effort was experienced. One important consideration in understanding individuals' evaluations of cognitive effort is the relative—versus absolute—nature of our subjective perceptions of magnitudes. That is, the perceived value (or unpleasantness) of a good, action, or experience depends strongly on the context in which it is evaluated (Hunter & Daw, 2021; Otto et al., 2022; Rangel & Clithero, 2012; Tversky & Simonson, 1993). Thus, an open question concerns whether our evaluations of cognitive demand—and the ensuing effort-based decisions informed by these evaluations—also occur in a relative versus absolute manner. Here, we sought to investigate whether our preferences for cognitive effort exertion—which presumably reflect our subjective evaluations of effort (Shenhav et al., 2017)—are indeed based on context-dependent valuations of effort cost.

It is worth noting that two opposing patterns of context effects have been previously observed. Under a so-called contrast effect, the subjective value of a given stimulus is biased away from its surrounding context. For instance, a rewarding option or course of action is increased in a low-value context (defined by small reward amounts) and is decreased in a

✉ A. Ross Otto
ross.otto@mcgill.ca

[1] Department of Psychology, McGill University, Montreal, Canada

high-value context (defined by large reward amounts; Bavard et al., 2018; Otto & Vassena, 2021; Palminteri et al., 2015). Similarly, in the aversive domain, individuals are willing to pay a greater price to be relieved of the same objective pain in a low-pain context than in a high-pain context (Vlaev et al., 2009). At the same time, other studies have revealed context effects whereby evaluations of stimuli are biased towards immediately surrounding stimuli—so-called assimilation effects. For example, faces of average attractiveness tend to be evaluated as more attractive when surrounded by highly attractive faces than when surrounded by other average faces (Geiselman et al., 1984). Similarly, consumer goods surrounded by moderately expensive goods tend to be judged as more expensive than when surrounded by inexpensive goods (Herr, 1989). In other words, assimilation effects occur when evaluations are pulled toward a reference point defined by a context, whereas contrast effects occur when evaluations are pulled away from a context. It is also worth noting that in some circumstances, the context surrounding an experience may not only affect evaluation of the experience but can also influence the individual's experience itself. For example, the educational psychology literature finds that that equally able students attending schools composed of above-average ability students exhibit a worse perception of their ability—and as a result, decreased academic performance—compared with students in low-average-ability schools (Marsh, 1987). This observed contrast effect suggests that context may also affect an individual's experience of effort itself, including their resultant performance, over and above the evaluation of effort.

Accordingly, we sought to explore the nature of possible context effects in choices concerning cognitive effort exertion. Following previous studies manipulating value context (Madan et al., 2021; Otto & Vassena, 2021), we devised an experiment whereby participants learned to associate unique stimuli with two different demand contexts. In our study, participants first completed an Association phase (see Fig. 1), where they learned to associate unique stimuli with different levels of a cognitively demanding 2-back working memory task (Buhle & Wager, 2010; Rischer et al., 2020) within two broader contexts: a low-demand context, defined by a sequence of low and medium levels of demand, and a high-demand context, consisting of medium and high levels of demand. Crucially, two distinct stimuli were associated with the same objective demand level across demand contexts (see Fig. 1E), resulting in four distinct stimuli to learn: low-demand in a low context (denoted 1), medium-demand in a low context (denoted 2LC), medium-demand in a high context (denoted 2HC), and high-demand in a high context (denoted 3).

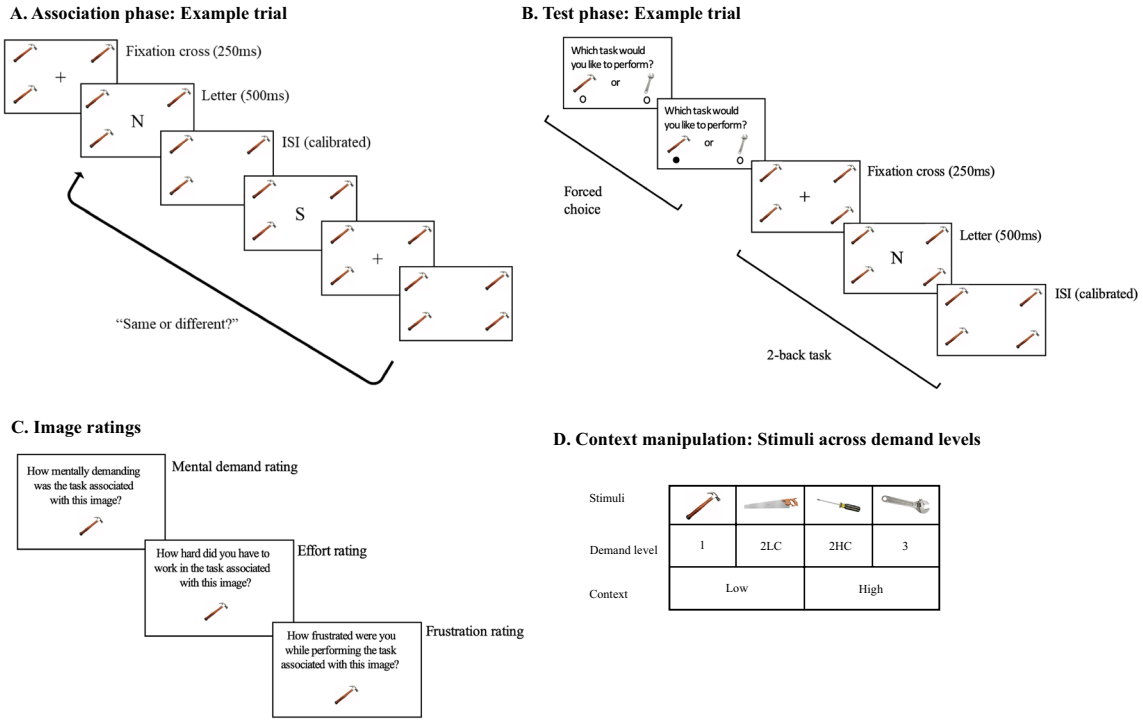

Fig. 1 **A** Example of an Association phase trial. During each trial of the Association phase, participants performed one trial of the 2-back task, after which they rated the mental demand, effort, and frustration associated with the demand level of the trial. **B** Example of a Test phase choice trial. Participants were first asked to choose one of two stimuli previously associated with demand levels. Following their choice, participants performed one trial of the 2-back task at their chosen demand level. **C** TLX ratings of to-be-learned stimuli. Between certain trials, participants rated the mental demand, effort, and frustration associated with each stimulus. **D** Four unique stimuli were paired with each demand level (1, 2LC, 2HC, 3) across two contexts (low and high)

In the Test phase, which followed the design of previous demand selection paradigms (Devine & Otto, 2022; Sayalı & Badre, 2019), participants then made decisions between pairs of stimuli, which allowed us to probe for context dependence in participants' learned valuations of the demand levels associated with each stimulus. Specifically, if valuations of cognitive demand levels are learned in an absolute (i.e., context-independent) fashion, participants should be indifferent when faced with the choice between the 2LC and 2HC stimuli, but if effort valuations are indeed shaped by the context in which they were learned, we should observe a reliable preference for one stimulus over the other.

While either pattern of context dependence—contrast or assimilation—could potentially manifest in participants' choices here, we had no strong prior hypotheses about which pattern we might observe. Under the assumption that participants prefer courses of action they evaluate as less cognitively demanding, under a contrast effect, the low-demand context would bias learned evaluations of the 2LC stimulus upward (i.e., would "pop out" of the surrounding context), which would manifest in a marked preference for the 2HC over the 2LC stimulus. In contrast, under an assimilation effect, we expected that participants would show a preference for the 2LC over the 2HC stimulus, as the surrounding low-demand context in which the 2LC option was learned would "pull in" its subjective value. Accordingly, in the present study, we sought to examine (1) whether individuals' evaluations of cognitive effort are indeed context-dependent and (2) the specific nature of this context dependence.

Finally, we examined participants' subjective effort ratings associated with each stimulus using the well-characterized TLX scale (Hart, 2006)—both "on-line" during the course of the Association phase and "off-line," immediately after the Test phases of the experiment (Fig. 1)—to probe for possible context effects in self-reported evaluations of cognitive demand.

Methods

Participants

We recruited 113 participants aged 18–70 years old (mean age = 21.88 years, 24 males) through McGill University. Participants were compensated at a rate of one course credit or $10 per hour. Participants provided informed consent in accordance with the McGill University Research Ethics Board (#398–0217). We excluded 14 participants with poor accuracy ($A < 0.75$; see below) on low-demand trials during the Association phase. The remaining sample consisted of 99 participants.

Procedure

First, participants completed a 2-back calibration task to determine the levels of difficulty to be used in the main task. Following this, participants were asked to complete a series of questionnaires (see below), and began the Association phase, where participants performed the 2-back task at three levels of difficulty across the high-demand and low-demand contexts and learned to association distinct stimuli with each context. In the Test phase of the main task, participants were asked to make decisions between the stimuli they learned about in the Association phase. The experiment lasted approximately two hours.

Calibration phase: 2-back task

Participants first performed a calibration task to adjust for individual differences in 2-back performance (Buhle & Wager, 2010). In this task, a series of letters appeared one at a time, and participants were asked to respond whether the current letter is the same or different as the letter presented two letters prior. Participants used the mouse buttons to indicate whether each letter is the same ("left") or different ("right"). A fixation cross was presented for 250 ms before each letter, after which each letter was presented for 500 ms (see Fig. 1B). We operationalized the demand level as the interstimulus interval (ISI; the time between the presentation of the letter and the fixation cross) such that decreasing the ISI typically engenders increased difficulty (i.e., lower accuracy). During the calibration, participants performed the 2-back task and a staircase technique for each demand level (low, medium, and high) was employed to identify three levels of demand suited to their individual performance, following our past work (Vogel et al., 2020). Participants performed three staircase procedures of 20 trials each to identify three ISIs corresponding to three target accuracy levels, quantified by the sensitivity measure A (where 1 corresponds to perfect performance, 0.5 corresponds to chance-level performance and 0 corresponds to all responses being incorrect; Zhang & Mueller, 2005). This staircase procedure established, for each participant, low-demand (1), medium-demand (2), and high-demand (3) ISIs which yielded A values of 1, 0.75, and 0.5, respectively. Each trial of the calibration phase had a fixed duration of 20 seconds, while the number of letters displayed per trial varied based on the ISI, such that longer ISIs entailed fewer letter stimulus presentations (and vice versa).

Association phase

After the 2-back calibration, participants completed the Association phase (Fig. 1B), which required participants to perform the 2-back task across two blocks, which defined the demand contexts. Specifically, in the

low-demand context (LC), participants performed the task at low and medium demand levels (1 and 2LC), and in high-demand context (HC), participants performed the task at medium and high levels demand (2LC and 3). The order of the low- and high-demand context blocks was counterbalanced across participants. Each block contained 32 trials, for a total of 64 trials during the Association phase. As with the calibration phase, each 2-back trial lasted 20 seconds, with the number of letters varying with the ISI.

During each trial, an arbitrary, affectively neutral image—drawn from a set of images of tools previously used in an associative learning study (Dunsmoor et al., 2014)—was displayed on-screen, each of which was associated with one of the three calibrated levels of 2-back difficulty. Crucially, two different stimuli were associated with the middle level of demand, across the two surrounding contexts, which yielded four distinct stimuli termed 1, 2LC (low context), 2HC (high context), and 3.

Subjective effort ratings

For a subset of 2-back trials during the Association phase, participants were prompted to provide ratings of subjective demand imposed by the most recently completed trial, for a total of 16 ratings. Specifically, participants rated the mental demand, effort, and frustration experienced during most recent 2-back trial, following the wording of the questionnaire items in the NASA Task Load Index (TLX; Hart & Staveland, 1988). Periodically during the Association phase, participants were asked to provide (on-line) TLX ratings of the mental demand, effort, and frustration associated with each image (see Fig. 1D). Finally, following choices in the Test phase, participants were required to complete another set of TLX ratings of images alone (which we termed off-line ratings) for a total of TLX ratings.

After the Association phase, participants made a series of forced choices in which they were required to choose between pairs of stimuli previously associated with demand levels (see Fig. 1C). This phase consisted of choices between four unique stimulus pairs: 1 versus 3, 1 versus 2LC, 2LC versus 2HC, and 2HC versus 3. Each choice pair was presented once, for a total of four choice trials. Following each choice, participants were required to complete a single trial of the 2-back task at the chosen demand level, lasting 20 seconds. After the four choice trials, participants provided a final set of TLX off-line ratings (mental demand, effort, frustration) for each stimulus.

Data analysis

We analyzed 2-back task performance (*A* and mean correct RTs) and subjective demand ratings using linear mixed effects regressions implemented with the *lme4* package for R (Bates et al., 2015). We included trial number as a predictor variable to account for the effect of elapsed time and estimated random intercepts over subject. We first estimated models predicting these outcomes as a function of objective demand level (1, 2, and 3), then estimated separate models directly comparing medium demand levels across low- versus high-demand context (2LC and 2HC). We tested the significance of demand level and context within each model using analyses of variance (ANOVAs) and performed pairwise comparisons using the *emmeans* package (Lenth et al., 2022). We assessed Test phase preferences using exact binomial tests and examined individual differences in preferences as a function of demand ratings and task accuracy using fixed-effects logistic regressions.

Results

2-back task performance

As expected, 2-back task accuracy across the Association phase significantly differed across the three objective demand levels, $F(2, 6630.6) = 780.81$, $p < 0.0001$, such that 2-back accuracy was significantly higher for low demand trials (mean sensitivity $A = 0.90$, $SD = 0.15$) than for medium demand trials ($A = 0.81$, $SD = 0.17$; $b = 0.09$, $SE = 0.005$, $p < 0.0001$) and for high demand trials ($A = 0.69$, $SD = 0.20$; $b = 0.21$, $SE = 0.005$, $p < 0.0001$). Additionally, task accuracy was significantly higher on medium demand trials than high demand trials ($b = 0.12$, $SE = 0.005$, $p < 0.0001$). Moreover, 2-back task accuracy on medium demand trials significantly differed across contexts, $F(1, 3067) = 40.95$, $p < 0.0001$, such that accuracy was lower on high-context (2HC) trials ($A = 0.83$, $SD = 0.18$) than on low-context (2LC) trials ($A = 0.79$, $SD = 0.15$; $b = -0.03$, $SE = 0.005$, $p < 0.0001$).

Turning to response times (RTs), we observed that correct RTs during the 2-back task significantly differed across demand levels, $F(2, 6234) = 6.30$, $p = 0.002$, such that RTs were significantly slower for low demand (RT = 315.82, $SD = 67.79$) than for medium demand trials (RT = 311.26, SD = 40.81; $b = 4.39$, $SE = 1.45$, $p = 0.007$), and RTs were significantly faster on high demand trials (RT = 314.83, $SD = 42.94$) than on medium demand trials ($b = -4.02$, $SE = 1.45$, $p = 0.016$). Correct RTs did not significantly differ between low and high demand trials ($b = 0.37$, $SE = 1.68$, $p = 0.97$). Finally, we observed that correct 2-back RTs on medium demand trials differed across contexts, $F(1,$

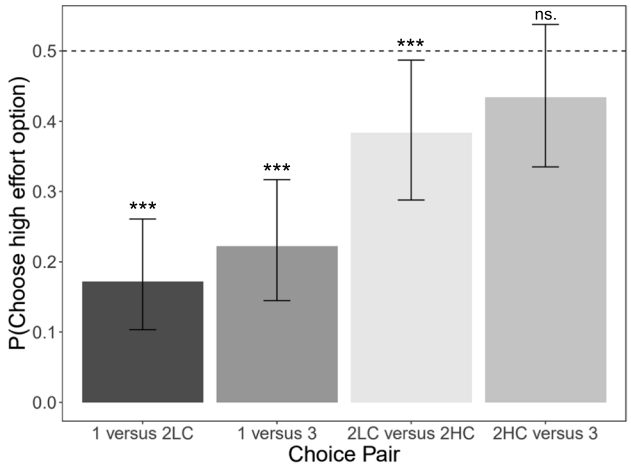

Fig. 2 The probability of choosing the higher-effort option across different pairings of stimuli associated with demand levels (Stimulus 1 versus 2LC, 1 versus 3, 2LC versus 2HC, 2HC versus 3). Error bars represent 95% confidence intervals. The dotted line represents chance level (0.5). Asterisks denote statistical significance in comparison to chance level (*** = significant with $p < .001$. ns = nonsignificant)

Table 1 Counts of the number of participants choosing each stimulus of the test phase

Choice pair	Low-effort choice	High-effort choice
1 versus 2LC	82	17
1 versus 3	77	22
2LC versus 2HC	61	38
2HC versus 3	56	43

3067) = 14.44, $p = 0.0002$, such that RTs were significantly faster on high-context (2HC) trials (RT = 308.42, $SD = 39.12$) than on low-context (2LC) trials (RT = 314.10, $SD = 42.44$; $b = -5.59$, $SE = 1.47$, $p = 0.0001$).

Demand preferences

Figure 2 and Table 1 depict participants' preferences for the higher-demand stimulus in the Test phase, across different option pairs (e.g., preference for choosing stimulus 3 over stimulus 1). We observed that, across stimulus pairs, participants generally exhibited a preference for the lower-demand option, replicating previous work (Devine & Otto, 2022; Kool et al., 2010; Sayalı & Badre, 2019). Specifically, participants were significantly less likely to choose demand Level 3 over demand Level 1, P(choose level 3) = 0.22; $p < 0.001$, and were less likely to choose demand Level 2LC over demand Level 1, P(choose level 2LC) = 0.17; $p < 0.001$. While participants numerically preferred demand Level 2HC over demand Level 3, P(choose level 3) = 0.43, this proportion was not significantly different from chance ($p = 0.23$).

Of central interest was participants' relative preference for the 2LC versus 2HC stimuli, which would indicate whether the demand context in which the stimuli—which were objectively identical in terms of associated demand level—would systematically shift evaluation of the demand associated with each stimulus, in turn engendering a reliable preference for one stimulus over the other. In the absence of a context effect, we expected participants to be indifferent between the two stimuli. Under a contrast effect, we expected participants to show a greater propensity for 2HC over 2LC. Alternatively, we expected participants to show a preference for 2LC over 2HC given an assimilation effect. Consistent with an assimilation effect, we observed that participants were significantly more likely to choose the 2LC stimulus over the 2HC stimulus, P(choose level 2HC) = 0.38; $p < 0.001$.

Subjective effort ratings

Figure 3 depicts both on-line (Association phase) and off-line (Test phase) TLX ratings for mental demand, effort, and frustration for the four stimuli during the Association. During both the Association and Test phases, we observed a significant effect of demand level on mental demand—association: $F(2, 1482) = 356.67$, $p < 0.0001$; test: $F(2, 294) = 98.11$, $p < 0.0001$; effort—association: $F(2, 1482) = 305.67$, $p < 0.0001$; test: $F(2, 294) = 100.26$, $p < 0.0001$; and frustration—association: $F(2, 1482) = 208.49$, $p < 0.0001$; test: $F(2, 294) = 67.40$, $p < 0.0001$—such that subjective demand ratings of mental demand, effort, and frustration increased with demand level, indicating that participants' self-reported demand evaluations tracked objective demand. Probing for potential context effects, we observed that in the Association phase, subjective ratings of mental demand, effort and frustration of medium demand significantly differed across contexts—mental demand: $F(1, 691) = 4.43$, $p = 0.04$; effort: $F(1, 691) = 5.94$, $p = 0.02$; frustration: $F(1, 691) = 9.54$, $p = 0.002$—such that subjective demand ratings were significantly greater for 2HC than 2LC stimuli in the Association phase (mental demand: $b = 0.47$, $SE = 0.23$, $p = 0.04$; effort: $b = 0.57$, $SE = 0.24$, $p = 0.02$; frustration: $b = 0.79$, $SE = 0.25$, $p = 0.002$) suggesting an assimilation effect, while we only observed a significant difference between 2HC and 2LC ratings for self-reported mental demand in the Test phase—mental demand: $F(1, 114.06) = 1.87$, $p = 0.017$, but not effort, $F(1, 113.90) = 2.86$, $p = 0.09$, or frustration, $F(1, 107.99) = 2.60$, $p = 0.11$.

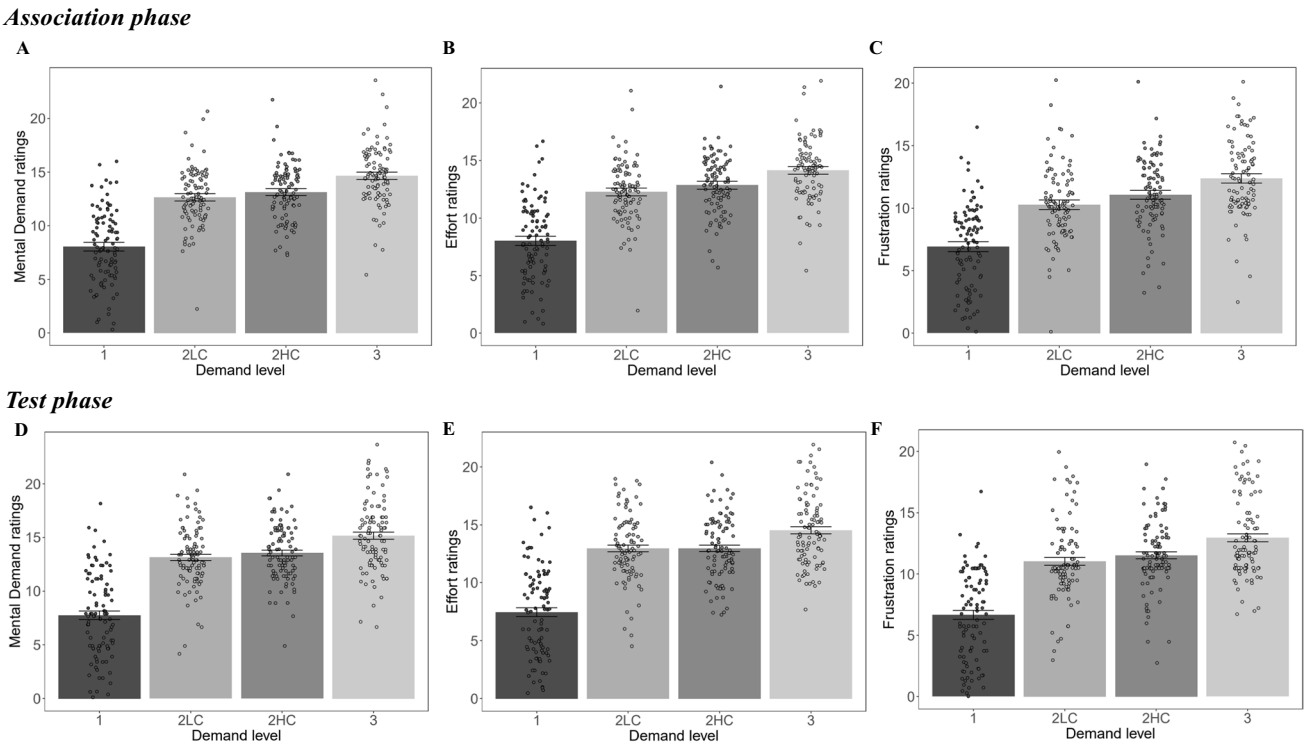

Fig. 3 Subjective ratings of mental demand, effort, and frustration during the association ("on-line"; **A–C**) and test phase ("off-line"; **D–F**). Error bars represent standard error of the mean

Analyses of individual differences

While we did not observe consistent context effects in participants' Test phase (off-line) stimulus ratings at the level of the entire sample, we could examine whether the considerable individual differences in context effects evident in off-line demand ratings (Fig. 4A) were related to individual-level context effects in preferences. In other words, we tested whether participants who rated the 2HC stimulus as more demanding than the 2LC stimulus were more likely to choose the 2LC stimulus over the 2HC stimulus. Figure 4B visualizes this predictive relationship, plotting

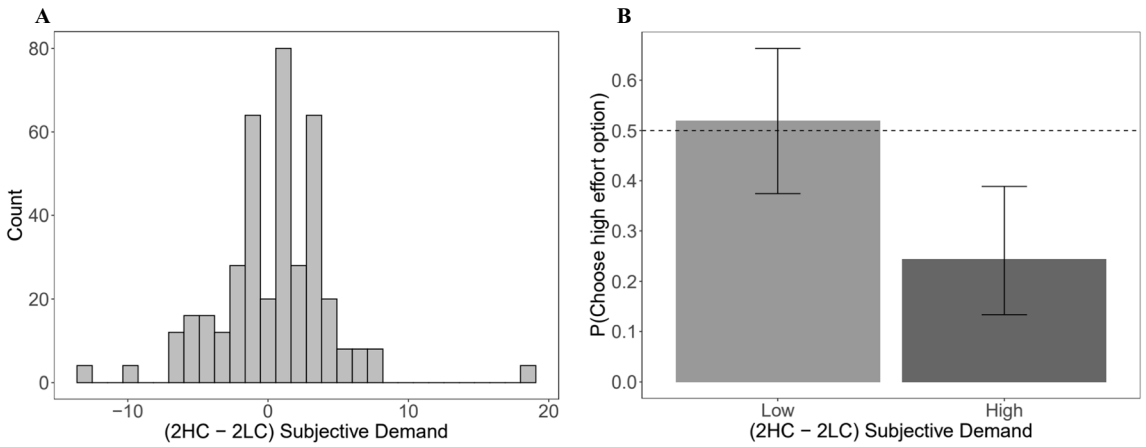

Fig. 4 **A** Histogram of differences in subjective ratings of demand (specifically, the "effort" rating) between the 2HC and 2LC stimulus. A more positive difference indicates a larger assimilation effect manifesting in off-line effort ratings. **B** Preference for the 2HC over the 2LC stimulus, as a function of the difference in subjective demand ratings between 2HC and 2LC (defined by a median split over the differences of composite demand ratings). Error bars represent confidence intervals. The dotted line represents chance level (0.5)

participants' preferences for the 2HC option over the 2LC separately for participants with small versus large difference in ratings between the 2HC and 2LC ratings (defined by a median split; larger values indicate a larger assimilation effect), which suggests participants were more likely to choose the 2LC stimulus if they rated higher subjective demand for 2HC stimulus than 2LC stimulus. Statistically, a series of separately estimated logistic regression models revealed that larger differences in mental demand ($b = -0.23$, $p = 0.002$), effort ($b = -0.25$, $p = 0.002$), and frustration ($b = -0.29$, $p < 0.001$) all significantly predicted a lower propensity to choose the 2HC stimulus over the 2LC stimulus. These observed individual differences are consistent with an assimilation effect, such that the more participant evaluated the high-demand context as more demanding, effortful, and frustrating than the low-demand context, the more likely they were to select the low-demand context (2LC) as opposed to the high-demand context (2HC) during the Test phase.

Similarly, we examined the extent to which individual differences in task performance between the low-demand context (2LC) and the high-demand context (2HC) were related to individual-level context effects in preferences. In other words, we tested whether participants who performed worse on high-demand trials (2HC) than on low-demand trials (2LC) during the Association phase were more likely to choose the 2LC stimulus over the 2HC stimulus in the Test phase. A logistic regression found that differences in accuracy between 2LC and 2HC trials did not significantly predict the propensity to choose one stimulus over the other during the Test phase ($b = 2.40$, $p = 0.28$), suggesting that participants were no more likely to choose the 2LC stimulus over the 2HC stimulus even if they previously performed more poorly on 2HC than 2LC trials.

Discussion

Our study sought to examine the extent to which evaluations of cognitive effort—evinced by demand preferences and subjective demand ratings—are context-dependent. To do this, participants performed a working-memory task at three different objective demand levels across low- and high-demand contexts, learning to associate distinct stimuli with each contextually bound demand level. When forced to choose between stimuli associated with objectively identically demand levels learned in different demand contexts, we found that participants markedly preferred a stimulus learned in a low-demand context over the (objectively identical) stimulus learned in a high-demand context, over and above generally demand-avoidant preferences. This pattern of preferences suggests that effort valuations are learned in a contextual fashion, and specifically, assimilate to the broader demand context in which the stimulus was learned.

Interestingly, we also found that participants exhibiting a larger assimilation effect in their subjective demand ratings—measured off-line, outside of demand contexts—also evidenced a larger assimilation effect in choices, manifesting in stronger preferences for the stimulus learned in a low-demand context. This observed predictive relationship suggests a concordance between contextual modulation of subjective evaluations and preferences, buttressing the idea that the observed demand preferences are based on subjective (and contextually bound) evaluations of demand. Further, and possibly mirroring the observed patterns of choice, participants rated this stimulus in the high-demand context (2HC) as more effortful than when in the low-demand context (2LC) during the Association phase.

It is worth noting that the high-demand context (2HC) was rated as more effortful and frustrating than the low-demand context (2LC) in on-line (i.e., during the Association phase) but not in off-line evaluations (during the Test phase). This discrepancy we observed between off-line ratings and preferences is evocative of the well-documented disconnect between self-reported valuations and preferences (Lichtenstein & Slovic, 1971). Indeed, while evaluations of stimuli may be more hypothetical and abstract, choices carry real consequences and therefore might reflect participants' true preferences about effort. Still, we found that the strength of context effect manifesting off-line ratings predicted individual-level context effects in demand preferences (Fig. 4B), suggesting that these off-line ratings evince, to some extent, the demand context surrounding the stimuli during learning.

However, the effects specifically observed in on-line ratings may be driven, in part, by sequential effects which often arise when individuals perform sequences of simple perceptual judgments. In particular, magnitude judgements are strongly influenced both by previously encountered stimuli, and individuals' history of responses (Donkin et al., 2015; Larsen & Norris, 2009). For example, loudness judgments have been demonstrated to be influenced by recent stimuli, such that a sound is judged as louder if it is directly preceded by a loud sound, and quieter if the loud sound occurred earlier in the series (Holland & Lockhead, 1968). Thus, an open question concerns the extent to which the assimilation effect observed in on-line effort ratings here is a pure reflection of participants' evaluations of demand associated with the stimuli (which are presumably contextually bound).

A further question raised by these results concerns the potential mechanism(s) that may underlie the observed assimilation effects. It is possible that cognitive fatigue may play a role in these contextually bound evaluations of stimuli, such that during the Association phase, the high-demand context may engender greater cognitive fatigue, resulting in the medium-demand (Level 2) task

being evaluated as more demanding in the high-demand versus the low-demand context (Lorist et al., 2005). Again, consistent with this idea, we observed that participants performed the 2-back task less accurately on 2HC trials than on (objectively identical) 2LC trials, possibly suggesting that fatigue engendered by the high-demand context impaired 2-back performance. At the same time, participants were faster to respond on these (again, identical) 2-back trials in the high-demand (2HC) versus low-demand (2LC) context, suggesting that participants may have strategically withheld effort on 2HC trials in the service of preserving performance on (more demanding) Level 3 trials and avoiding fatigue (Hockey, 2013; Matthews et al., 2023). Alternatively, performing the high-demand task may induce a negative affective state (e.g., frustration), resulting in the high-demand context being evaluated more negatively, and therefore, as more effortful (Chen et al., 2023; Devine et al., 2023). Finally, the observed context effects may be the result of an anchoring process (Seymour & McClure, 2008). Consistent with evaluations of facial attractiveness (Geiselman et al., 1984), consumers' prices evaluations (Cunha & Shulman, 2011) and categorization of ambiguous stimuli (Herr, 1989), the demand contexts established in the present study may anchor evaluations of cognitive effort such that subjective value is pulled upward in a high-demand context and pulled downward in a low-demand context.

As noted above, participants performed less accurately on 2HC trials than 2LC trials during the Association phase, suggesting that the demand context not only affected the value of the medium-demand level but also how it was experienced by the participants, such that their performance on (objectively identical) demands level varied as a function of surrounding context. While the medium demand level was identical within the low- and high-demand contexts in our design, these surrounding contexts appeared to affect participants' experience of—and consequently, performance—on these trials. In this sense, the context effect we observed diverges from "pure" context effects observed in economic settings whereby the mere presence of items in the surrounding context is thought to bias evaluation of a good or action. Our results rather suggest that these demand contexts produce temporally sustained effects on task performance, above and beyond the effects of the presence of context-defining stimuli. While our results suggest a context effect as the surrounding demand context biased participants' evaluations (and choices regarding) of objectively identical demand levels, they outline a fundamental challenge in isolating experienced effort from other factors that may influence task performance (Fleming et al., 2023). Accordingly, future work should attempt to disentangle the mere presence of an option and these temporally sustained effects of the option on subjective effort.

Although we had no strong a priori hypothesis about the nature of the observed context effect, it is noteworthy that effort evaluations manifest an assimilation rather than a contrast effect (Bavard et al., 2018; Otto & Vassena, 2021; Vlaev et al., 2009). One possible explanation for the assimilation (versus contrast) effect we observed may be the moderate spread of 2-back difficulty levels experienced within each context. In particular, while ambiguous values tend to assimilate to contexts with moderate ranges, contexts defined by more extreme ranges of values tend to result in contrast effects (Herr, 1989). On this view, owing to the relatively similar demand levels evoked by our ISI manipulation—resulting in observed A (discriminability) levels ranging from 0.69 to 0.90—one might expect that evaluations of the 2LC and 2HC stimuli would assimilate to their respective contexts. Previously, we have observed contrast effects in reward incentive evaluation (rather than cognitive demand evaluation) examining responsivity to 10-cent incentives across low-reward (1 and 10 cent) and high-reward (10 and 19 cent) incentives (Otto & Vassena, 2021), which could be interpreted as more extreme differences in reward levels. Accordingly, one open question concerns whether more extreme differences in cognitive demand (than used in the present study) might produce contrast, as opposed to assimilation effects. Future empirical work, ideally using the present task structure, would be helpful to examine this possibility. Finally, and echoing the potential fatigue explanation proffered above, Martin et al. (1990) have suggested that the processes involved in contrast effects are more cognitively effortful than those involved assimilation. Here, it may be the case that the processing load exacted by the 2-back task in the Association phase displaces processing—which would otherwise give rise to a contrast effect—reverting the evaluative process to one based on assimilation.

The present study provides an initial investigation of the context-dependent nature of effort evaluations highlighting that, like rewards (Bavard et al., 2018; Palminteri et al., 2015), the context in which cognitive demand is experienced can systematically shift our demand evaluations. In turn, such context dependence might hold the consequence that our decisions to expend cognitive effort might be informed by exaggerated subjective effort costs associated with a particular course of action if it is encountered within a high-demand context, and at the same time, suggests that low-demand contexts could be used to reduce the aversiveness of cognitive effort.

Funding This work was supported by grants from the Engineering Research Council of Canada, the Fonds de Recherche du Québec–Nature et Technologies, and the Canada Foundation for Innovation (to ARO) and a Fellowship from the Fonds de Recherche du Québec–Nature et technologie (to SD).

Data availability All raw data pertaining to this study can be accessed via the Open Science Framework (https://osf.io/3w9at/).

Code availability Not applicable.

Declarations

Conflicts of interest Not applicable.

Ethics approval This procedure was approved by the McGill Research Ethics Board (REB #137–0816).

Consent to participate All participants gave informed consent prior to testing.

Consent for publication Not Applicable.

Open practices statement Data or materials for the experiments are available upon request and, upon acceptance, all raw data will be made publicly available on the Open Science Foundation (www.osf.io) website. This experiment was not preregistered.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bavard, S., Lebreton, M., Khamassi, M., Coricelli, G., & Palminteri, S. (2018). Reference-point centering and range-adaptation enhance human reinforcement learning at the cost of irrational preferences. *Nature Communications*, *9*(1), Article 4503. https://doi.org/10.1038/s41467-018-06781-2

Buhle, J., & Wager, T. D. (2010). Performance-dependent inhibition of pain by an executive working memory task. *Pain, 149*(1), 19–26. https://doi.org/10.1016/j.pain.2009.10.027

Chen, J., Huys, Q. J., Stringaris, A., & Nielson, D. M. (2023). On the misery of cognitive effort. *PsyArXiv Preprint*. https://doi.org/10.31234/osf.io/r9e4j

Cunha, M., & Shulman, J. D. (2011). Assimilation and contrast in price evaluations. *Journal of Consumer Research, 37*(5), 822–835. https://doi.org/10.1086/656060

Devine, S., & Otto, A. R. (2022). Information about task progress modulates cognitive demand avoidance. *Cognition*, *225*, Article 105107. https://doi.org/10.1016/j.cognition.2022.105107

Devine, S., Vassena, E., & Otto, A. R. (2023). More than a feeling: Physiological measures of affect index the integration of effort costs and rewards during anticipatory effort evaluation. *Cognitive, Affective, & Behavioral Neuroscience, 23*(4), 1129–1140. https://doi.org/10.3758/s13415-023-01095-3

Donkin, C., Rae, B., Heathcote, A., & Brown, S. D. (2015). Why is accurately labeling simple magnitudes so hard? A past, present, and future look at simple perceptual judgment. In J. R. Busemeyer (Ed.), *The Oxford handbook of computational and mathematical psychology*. Oxford University Press.

Dunsmoor, J. E., Kragel, P. A., Martin, A., & La Bar, K. S. (2014). Aversive learning modulates cortical representations of object categories. *Cerebral Cortex, 24*(11), 2859–2872. https://doi.org/10.1093/cercor/bht138

Fleming, H., Robinson, O. J., & Roiser, J. P. (2023). Measuring cognitive effort without difficulty. *Cognitive, Affective, & Behavioral Neuroscience, 23*(2), 290–305. https://doi.org/10.3758/s13415-023-01065-9

Geiselman, R. E., Haight, N. A., & Kimata, L. G. (1984). Context effects on the perceived physical attractiveness of faces. *Journal of Experimental Social Psychology, 20*, 409–424.

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139–183). North Holland.

Herr, P. M. (1989). Priming price: Prior knowledge and context effects. *Journal of Consumer Research, 16*(1), 67–75. https://doi.org/10.1086/209194

Hockey, R. (2013). The psychology of fatigue: Work, effort and control. *Cambridge University Press.*. https://doi.org/10.1017/CBO9781139015394

Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception & Psychophysics, 3*(6), 409–414. https://doi.org/10.3758/BF03205747

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. Appleton-Century.

Hunter, L. E., & Daw, N. D. (2021). Context-sensitive valuation and learning. *Current Opinion in Behavioral Sciences, 41*, 122–127. https://doi.org/10.1016/j.cobeha.2021.05.001

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General, 139*(4), 665–682. https://doi.org/10.1037/a0020198

Larsen, J. T., & Norris, J. I. (2009). A facial electromyographic investigation of affective contrast. *Psychophysiology, 46*(4), 831–842. https://doi.org/10.1111/j.1469-8986.2009.00820.x

Lenth, R. V. (2022). Emmeans: Estimated Marginal Means, Aka Least-Squares Means. https://cran.r-project.org/package=emmeans

Lichtenstein, S., & Slovic, P. (1971). Reversals of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89*(1), 46–55.

Lorist, M. M., Boksem, M. A., & Ridderinkhof, K. R. (2005). Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cognitive brain research, 24*(2), 199–205. https://doi.org/10.1016/j.cogbrainres.2005.01.018

Madan, C. R., Spetch, M. L., Machado, F. M. D. S., Mason, A., & Ludvig, E. A. (2021). *Encoding Context Determines Risky Choice, 32*(5), 743–754.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79*(3), 280–295.

Martin, L. L., Seta, J. J., & Crelia, R. A. (1990). Assimilation and contrast as a function of people's willingness and ability to expend effort in forming an impression. *Journal of Personality and Social Psychology, 59*(1), 27–37. https://doi.org/10.1037/0022-3514.59.1.27

Matthews, J., Pisauro, M. A., Jurgelis, M., Müller, T., Vassena, E., Chong, T. T.-J., & Apps, M. A. J. (2023). Computational mechanisms underlying the dynamics of physical and cognitive fatigue. *Cognition*, *240*, Article 105603. https://doi.org/10.1016/j.cognition.2023.105603

Otto, A. R., & Vassena, E. (2021). It's all relative: Reward-induced cognitive control modulation depends on context. *Journal of Experimental Psychology: General, 150*(2), 306–313. https://doi.org/10.1037/xge0000842

Otto, A. R., Devine, S., Schulz, E., Bornstein, A. M., & Louie, K. (2022). Context-dependent choice and evaluation in real-world consumer behavior. *Scientific Reports, 12*(1), Article 1. https://doi.org/10.1038/s41598-022-22416-5

Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment

learning. *Nature Communications*, *6*(1), Article 8096. https://doi.org/10.1038/ncomms9096

Rangel, A., & Clithero, J. A. (2012). Value normalization in decision making: Theory and evidence. *Current Opinion in Neurobiology, 22*(6), 970–981. https://doi.org/10.1016/j.conb.2012.07.011

Rischer, K. M., González-Roldán, A. M., Montoya, P., Gigl, S., Anton, F., & Meulen, M. (2020). Distraction from pain: The role of selective attention and pain catastrophizing. *European Journal of Pain, 24*(10), 1880–1891. https://doi.org/10.1002/ejp.1634

Sayalı, C., & Badre, D. (2019). Neural systems of cognitive demand avoidance. *Neuropsychologia*, *123*(June 2018), 41–54. https://doi.org/10.1016/j.neuropsychologia.2018.06.016

Seymour, B., & McClure, S. M. (2008). Anchors scales and the relative coding of value in the brain. *Current Opinion in Neurobiology, 18*(2), 173–178. https://doi.org/10.1016/j.conb.2008.07.010

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience, 40*, 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Strobel, A., Wieder, G., Paulus, P. C., Ott, F., Pannasch, S., Kiebel, S. J., & Kührt, C. (2020). Dispositional cognitive effort investment and behavioral demand avoidance: Are they related? *PLOS ONE*, *15*(10), Article e0239817. https://doi.org/10.1371/journal.pone.0239817

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science, 39*(10), 1179–1189.

Vermeylen, L., Braem, S., Notebaert, W., & Ruitenberg, M. F. L. (2022). The subjective evaluation of task switch cues is related to voluntary task switching. *Cognition*, *224*, Article 105063. https://doi.org/10.1016/j.cognition.2022.105063

Vlaev, I., Seymour, B., Dolan, R. J., & Chater, N. (2009). The price of pain and the value of suffering. *Psychological Science, 20*(3), 309–3179.

Vogel, T. A., Savelson, Z. M., Otto, A. R., & Roy, M. (2020). Forced choices reveal a trade-off between cognitive effort and physical pain. *eLife*, *9*(November), 1–18. https://doi.org/10.7554/eLife.59410

Westbrook, A., Kester, D., & Braver, T. S. (2013). What is the subjective cost of cognitive effort? Load, trait, and aging effects revealed by economic preference. *PLoS ONE, 8*(7), 1–8. https://doi.org/10.1371/journal.pone.0068210

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika, 70*(1), 203–212. https://doi.org/10.1007/s11336-003-1119-8