PENULTIMATE DRAFT, for *WIREs Cognitive Science*

Conscious cognitive effort in cognitive control

Joshua Shepherd (Carleton University & Universität de Barcelona)

ABSTRACT

Cognitive effort is thought to be familiar in everyday life, ubiquitous across multiple variations of task and circumstance, and integral to cost/benefit computations that are themselves central to the proper functioning of cognitive control. In particular, cognitive effort is thought to be closely related to the assessment of cognitive control's costs. I argue here that the construct of cognitive effort, as it is deployed in cognitive psychology and neuroscience, is problematically unclear. The result is that talk of cognitive effort may paper over significant disagreement regarding the nature of cognitive effort, and its key functions for cognitive control. I highlight key points of disagreement, and several open questions regarding what causes cognitive effort, what cognitive effort represents, cognitive effort's relationship to action, and cognitive effort's relationship to consciousness. I also suggest that pluralism about cognitive effort – that cognitive effort may manifest as a range of intentional or non-intentional actions the function of which is to promote greater success at paradigmatic cognitive control tasks – may be a fruitful and irenic way to conceive of cognitive effort. Finally, I suggest that recent trends in work on cognitive control suggests that we might fruitfully conceive of cognitive effort as one key node in a complex network of mental value, and that studying this complex network may illuminate the nature of cognitive control, and the role of consciousness in cognitive control's proper functioning.

1 Introduction

Cognitive control is commonly understood as the set of processes or capacities that together drive flexible, often goal-directed, often adaptive mental processing (Botvinick et al. 2001, Rouault and Koechlin 2018, Badre and Nee 2018). Mainstream theorizing takes cognitive control to be at work in operations like selective information retrieval, maintenance of task-relevant information, monitoring of working memory contents, task-relevant inhibition of habitual responses, construction of task sets, and switching between task sets. These operations are studied in inhibition paradigms like those that use go/no-go tasks, or tasks that involve Stroop stimuli, in rule shifting paradigms like those that use the Wisconsin Card Sorting Task, in action planning paradigms like those that use the Tower of London task, and in information maintenance paradigms like those that use N-back tasks (see Miyake et al. 2000, Gratton et al. 2018).

The deployment of cognitive control will typically involve the selection of a mental operation (or a package of mental operations) from a broader set of options, and may also involve the monitoring of the progress of these operations. The exact nature of these operations, sometimes discussed as the generation of 'control signals,' is often left open. Instead, more work has focused upon how we select which control signals to deploy. This work is guided by a double move that first conceptualizes cognitive control – very plausibly – as resource limited. Given limitations, a plausible assumption is that a cognitive control system that approximates practical wisdom by efficiently distributing resources, or efficiently generating packages of control operations, will do better. The second part of the move is to conceptualize the decision process that leads to resource distribution using cost-benefit computations (Kool et al. 2017, Kool and Botvinick 2018).

According to the Expected Value of Control framework (Shenhav et al. 2013, Botvinick and Braver 2015, Shenhav et al. 2017), for example, the cognitive control system monitors the identity and intensity of available control signals, and seeks to measure the likelihood and magnitude of reward given a specification of a control signal along these two dimensions. In addition to the probability of reward, various costs have to be taken into account. For example, there is a basic limit on working memory capacity, such that entrance into working memory and exit from working memory may require gating policies specified in terms of cost-benefit computations (see, e.g., Braver and Cohen 2000, Dayan 2012, Chatham and Badre 2015). In the end, a good cognitive control system is one that approaches optimality at taking the actions most likely to generate positive reward/cost balances.

It is here, at least on this telling of the story, that 'cognitive effort' (sometimes called 'mental effort') enters in. Cognitive effort is thought to be familiar in everyday life, ubiquitous across multiple variations of task and circumstance, and integral to cost/benefit computations that are themselves central to the proper functioning of cognitive control (Kurzban et al. 2013, Kurzban 2016, Shenhav et al. 2017, Kool and Botvinick 2018). In particular, cognitive effort is thought to be closely related to the assessment of cognitive control's costs.

At this point a potential complication arises, because the term 'effort' has a long history in psychology and physiology, and it is unclear whether research on effort always has the same referent across disciplines like exercise and motor physiology, sport psychology, and cognitive neuroscience. It is possible, for example, to distinguish between cognitive effort and physical effort. But it remains an open question just how cognitive and physical effort may interact with and influence each other (Smit et al. 2005, Schmidt et al. 2012).[1] In what follows, my discussion targets only mental effort and cognitive effort as these notions arise in the cognitive control and decision-making literatures.

I argue here that the construct of cognitive effort, as it is deployed in cognitive psychology and neuroscience, is problematically unclear. The result is that talk of cognitive effort may paper over significant disagreement regarding the nature of cognitive effort, and its key functions for cognitive control. In sections two through five I discuss a range of important questions and choice points facing any attempt to account for cognitive effort (see Figure 1). In section two I discuss different characterizations of the nature of cognitive effort. In section three I discuss questions regarding what cognitive effort represents. In section four I discuss the relationship

[1] Richter et al. (2016) offer a good discussion of effort from the perspective of motivational intensity theory; James Steele offers a good and thorough review of work on effort, and perception of effort, across many disciplines (this review is on the arxiv, and this journal does not permit citations to the arxiv); Massin (2017) offers a good discussion of different accounts of effort across many disciplines, arguing for a 'force-based' account; Juan Pablo Bermúdez and Olivier Massin have work in progress that offers a nice philosophical overview of effort across many disciplines.

between cognitive effort, action, control, and automaticity. In section five I discuss the relationship between cognitive effort and consciousness. In section six I discuss how to proceed in the face of so much disagreement and so many open questions regarding cognitive effort. And in section seven I argue that recent trends in work on cognitive control suggests that we might fruitfully conceive of cognitive effort as one key node in a complex network of mental value, and that studying this complex network may illuminate the nature of cognitive control, and the role of consciousness in cognitive control's proper functioning.

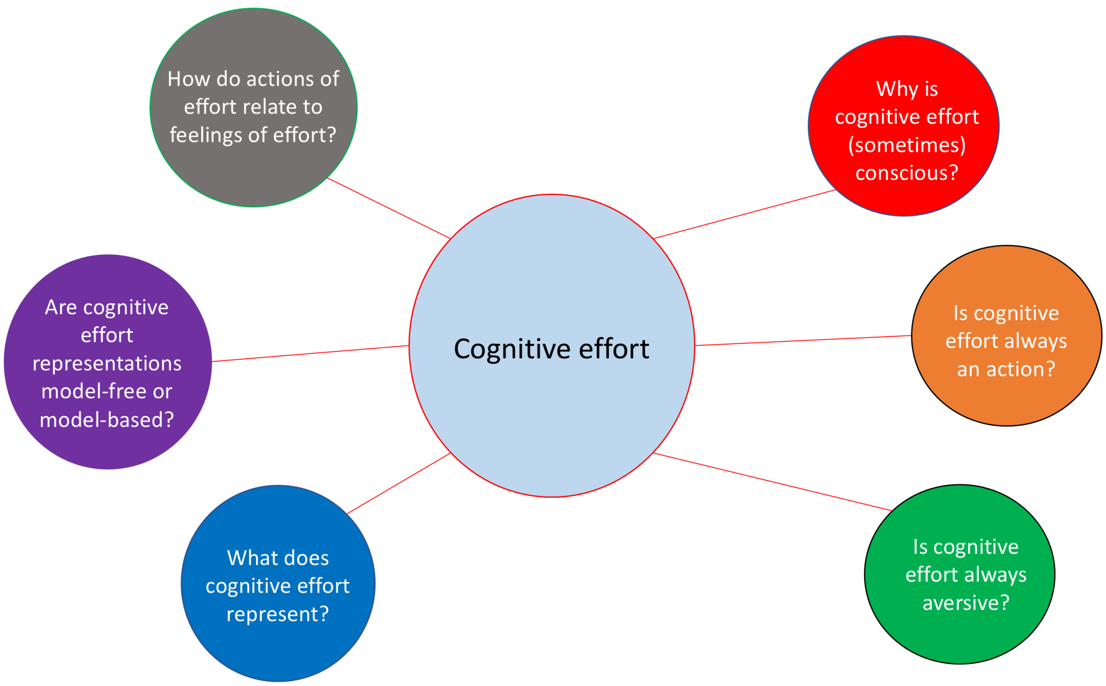

Figure 1. Choice points for a theory of cognitive effort.

2 The nature of cognitive effort

What is cognitive effort? On the one hand, there is a sense that cognitive effort's nature is fairly obvious. So, Shenhav et al. (2017) begin an influential recent paper by stating that 'Cognitive effort is among the most familiar and intuitive fixtures of mental life' (100). On the other hand, a close look at characterizations of the construct of cognitive effort reveals – beyond the general thought that cognitive effort plays a causal role in explanations of cognitive control performance – significant disagreement. And this disagreement tracks fault lines that are philosophically and theoretically important. The discussion in this section attempts to uncover these fault lines.

Shenhav et al. (2017) characterize cognitive effort as the thing that mediates between '(a) the characteristics of a target task and the subject's available information-processing capacity and (b) the fidelity of the information-processing operations actually performed, as reflected in task performance' (100). As they have it, then, cognitive effort is the 'set of intervening processes' that explains why, given the agent's capacity for performing at level 10 (say), the agent performed at level 8 (or 6 or whatever). Inzlicht et al. (2018) say something similar – cognitive effort is 'the process that mediates between' potential performance and actual performance (2018, 338).

These characterizations suggest that, no matter how we characterize cognitive control performance determinants (the set of intervening processes) at a finer grain, cognitive effort constitutes the collection. Cognitive effort is the entire set of performance determinants. But there is reason to doubt that this is the best way to think of cognitive effort. One reason is that performance seems determined, in part, not just by effort, but by the fidelity or appropriateness of the control operations the agent selects.[2] Think, e.g., of inhibiting instead of monitoring, or switching task at the right time instead of too late. One might invest a great amount of effort on the wrong control operation, leading to poor performance. And this indicates that the determination of performance is not simply a matter of investing effort. A second reason stems from reflection on easy tasks. There may be tasks that are relatively simple to perform, such that differences in effort investment provides no added benefit in terms of performance.

These issues raise interesting questions regarding the relationship between effort investment and performance. These questions include: why the deployment of cognitive effort enhances performance (see Khachouf et al. 2017), whether the deployment of cognitive effort is sometimes ineffective, and whether there are better and worse ways to deploy cognitive effort.

Shenhav et al.'s characterization of cognitive effort suggests that cognitive effort is to be understood as relative to the agent's available information-processing capacity. If so, then we would most plausibly characterize the expenditure of effort in terms of *amounts*. But then one wants to ask: amounts of what? A natural option is to understand amounts of effort in terms of the ratio of information-processing capacity deployed. But a separate option is that we should understand amounts of effort in terms of the intensity of the control signals deployed. And one way to characterize intensity here appeals to force-based accounts of effort (accounts often more popular in discussions of physical or bodily effort). Massin (2017), for example, has argued that effort should be understood as a voluntary (or intentional) action of 'exerting some force so as to produce some desired outcome' (243). As Massin recognizes, how to characterize mental or cognitive force in a way that allows further explication of a force-based account of cognitive effort remains open.

A distinct idea is offered by Székeley and Michael (2021). They argue that we should think of cognitive effort as mediating between the characteristics of the task and the agent's information-processing capacity, and 'the flexible adjustment of information-processing to optimise performance' (898). Here cognitive effort promotes the flexible adjustment of information-processing, but it is not directly explanatory of performance; cognitive effort mediates between capacity and flexibility, and the impact of effort on performance is only directed towards performance optimization. This leaves room for cases in which flexible adjustment fails to optimize.

---

[2] Shenhav et al.'s characterization does mention the idea of fidelity, but it is unclear how to square this idea with the idea that investing effort also has to do with investing some amount of information-processing resources.

We have, then, at least three options for characterizing cognitive effort. First, cognitive effort may be a dedication of some amount of information-processing capacity. Second, cognitive effort may be the level of intensity (or force) with which some control signal (or package of signals) is deployed. Third, cognitive effort may be the adjustment of information-processing, when the adjustments are meant to optimize performance. On the first two options, cognitive effort is always present in cognitive control. For there will always be some amount of information-processing, or some level of intensity, deployed. On the third option, cognitive effort may be absent in some tasks, since agents may not always adjust information-processing, and may not always seek to optimize performance.

3 What cognitive effort represents

Questions about the nature of cognitive effort are related to, and should be informed by, accounts of how cognitive effort arises, and what effort represents. How cognitive effort arises, and what cognitive effort represents, are of course also closely related. On a promising approach to representation (Shea 2018), for example, the content of a representation is fixed relative to the task functions the representation helps explain, and 'task functions arise as a result of some stabilizing process' (217) – stabilizing processes are processes like learning, or natural selection. On such an approach, the causal explanation for the appearance of cognitive effort in some task, and the functions cognitive effort is supposed to help perform, will provide information regarding the content of cognitive effort.

The current literature contains a number of subtly but importantly different ideas regarding the causes of, and possible functions of, cognitive effort. Kurzban et al. (2013) characterize mental effort as arising from the monitoring of opportunity costs of continuing the *current*, as measured against nearby, tasks. A chief function of mental effort, then, is to motivate shifts of attention away from the current task, towards more rewarding possibilities. This proposal seems to conceptualize cognitive effort as arising from model-based computations of opportunity cost, and suggests that what effort represents is the opportunity cost of continuing in the task. Carruthers and Williams (2022) propose that doing so would often be too computationally costly. They argue that cognitive effort is generated by model-free (and analog-magnitude) signals of executive engagement in a task – 'evolution has provided a sort of summary estimation of the opportunity costs, coded into a default value for mental effort' (12). On this proposal, cognitive effort seems to represent executive engagement.

Some favor a picture on which mental effort's function is built in to a neuroeconomic model that specifies cognitive effort as a cost in labor/leisure trade-off decisions (Kool and Botvinick 2014). As such, the cost of cognitive effort may vary according to context, with effort costing more, so to speak, 'when one is already working hard than when one is hardly working' (139). On this proposal, cognitive effort may be thought to represent the cost of increased labor, which is subtly distinct from a representation of opportunity costs. One's labor might be costly when compared to leisure, even if the opportunity cost of the labor is low compared to the cessation of work.

A very different idea is proposed by Lukitsch (2020), who ties the function of effort more closely to a model of action control, on which the sense of effort signals an uncertain trajectory, motivating the deployment of greater control. Here, cognitive effort seems to represent uncertainty of success.

The proliferation of such ideas in the literature signals difficulties for our attempts to give an account of what cognitive effort represents. One difficulty is simply that of isolating cognitive effort from the representations that may interact with cognitive effort to give fine-grained explanations of cognitive control operations, as well as the many causal factors that may be relevant to the generation of cognitive effort in different contexts. A second difficulty concerns the flexibility of representations of cognitive effort – how closely cognitive effort tracks things like opportunity cost, executive engagement, and contextual factors that carry information about these. Recent advances in computational modelling, and in methods such as representational similarity analysis, may provide traction for addressing these difficulties (see Freund et al. 2021).

4 Cognitive effort's relationship to action

A separate issue concerns the relationship of cognitive effort to intentional action – that is, the goal-directed causation of mental or physical activity by representational states like goals or intentions (Shepherd 2021a). Some theorists characterize cognitive effort as a 'subjective intensification of mental activity' (Kool and Botvinick 2018, 899, drawing on Inzlicht et al. 2018). This characterization suggests that cognitive effort is an action that subjects perform. Some recent proposals are specific about this. Massin (2017), for example, argues that effort is the intentional exertion of a force – 'effort, on a force-based account, contains at least two aspects: a force and a telos, that is, the force exerted by the agent, and the goal that the agent thereby pursues' (243). And in a recent paper linking cognitive control and self-control, Sripada (2021) proposes we think of these control signals as basic (mental) actions – a term from action theory indicating relatively primitive but intentional actions we perform without performing any other intentional action.

One question for this proposal, moving forward, is how agents might learn to combine and sequence basic mental actions to achieve sophisticated goals (see Chiu and Egner 2017). What kinds of mental flexibility are available to human agents (on which, see Khachouf et al. 2017), and what kind of constraints and limitations do we face? A second question is whether it is possible to distinguish between the intentional (or controlled) deployment of cognitive control, and the non-intentional – in what circumstances, if any, might deployment of cognitive control become relatively automatized (see Cushman and Morris 2015)? A third question concerns the kinds of control signals available. What kinds of control signals are available for the guidance of attention, memory, or thought? Are these signals are relatively fixed, or is it possible to learn new ways to guide thought over time?

Conceiving of the exertion of cognitive effort as an action has a long lineage. William James, for example, listed 'making a mental effort' alongside other mental actions like attending, assenting, and negating (James 1890, 287). Interestingly, if exerting cognitive effort is an action, one might expect greater flexibility in its deployment than if cognitive effort were only a set of computationally driven processes, or if cognitive effort were a conscious feeling. For the execution of action, on the common understanding, is to some extent under the subject's control (Shepherd 2021a). But of course some actions occur habitually or automatically, and recent work indicates that automaticity and control might be thought of as bound up together in the same action (Wu 2013, Pacherie and Mylopoulos 2021, Shepherd 2021b). How are we to distinguish between the intentional (or controlled) deployment of cognitive effort, and the non-intentional? To what extent might the selection of cognitive control operations, including decisions about effort investment, proceed automatically (see Cushman and Morris 2015, Chiu and Egner 2017)?

Consider an analogy with blinking, or with breathing. These operations often proceed independently of any decision made by the agent to blink or to breathe. But agents can take control of these processes (to some degree) – choosing to (not) blink or to (not) breathe in certain ways or for certain temporal durations. Perhaps actions of effort investment are like this. If so, an important question surrounds the limits of control with respect to actions of effort investment. How might agents learn to combine and sequence basic mental actions to achieve sophisticated goals (see Chiu and Egner 2017)? What kinds of mental flexibility are available to human agents (on which, see Khachouf et al. 2017), and what kind of constraints and limitations do we face?

5 Cognitive effort and consciousness

Many suggest that cognitive effort either is, or is typically associated with, a certain kind of conscious experience. Kurzban et al. (2013), for example, posit that cognitive effort is a felt sensation (signalling opportunity cost), that may impact downstream processing. And André et al. posit that 'effort is a feeling that emerges in awareness during effortful tasks' (2). Of course, effort-as-(conscious) feeling is not logically inconsistent with effort-as-performance-determinant. It might be that the processes that determine cognitive control performance are largely driven by the neural realizers of feelings of effort. But other lines of research posit other types of conscious feelings – anxiety, fatigue, boredom (Agrawal et al. 2022) – that are also supposed to impact cognitive control performance. How might cognitive effort relate to these other feelings? I consider this question further below.

Some draw a distinction between cognitive effort considered as something like an action related to information processing, and sensations or perceptions of effort (see, e.g., Hermann and Johnsrude 2020). Often, the relationship between cognitive effort and consciousness is thought to occur with respect to sensations or perceptions of effort. But it is possible that actions of effort investment have their own phenomenology, distinct from that of sensations or perceptions of effort (Kriegel 2015, Shepherd 2017).

The relationship between consciousness, effort-as-action, and effort-as-feeling remains unclear. Here are three relatively plausible options. First, effort-as-action might reliably cause feelings of effort. If so, one question concerns breakdowns in this causal pathway – when might actions of effort investment fail to produce feelings of effort? Is it possible to experience illusory feelings of effort? Second, one might argue that effort-as-action is identical with effort-as-feeling. Shepherd (2016), for example, argues that 'the neural realizers of experiences of trying (that is, experiences of directing effort towards the satisfaction of an intention) are not distinct from the neural realizers of actual trying (that is, actual effort directed towards the satisfaction of an intention)' (2016, 419). Third, effort-as-action might be related to effort-as-feeling via a mediated route. This seems to be the view of many in the cognitive control literature, who hold that feelings of effort are produced by computations that may take investment of resources as an input, but that also take on board additional factors related to the opportunity costs of performing a particular task (Kurzban et al. 2013, Székeley and Michael 2021).

These options are not mutually exclusive, if we posit different feelings associated with each of them. Neither, though, are choices between these options theoretically idle. For if feelings of effort are distinct from, but reliably associated with, actions of effort, then it is plausible that both items will be important for understanding performance on cognitive control tasks. The feeling of effort could, for example, influence the level of effort one maintains over the course of a task.

A further issue surrounds the conscious aspect of cognitive effort. Cognitive effort is discussed, almost always, as something of which agents are aware – I have been unable to find mention of unconscious cognitive (or mental) effort in the literature. Additionally, the exact relationship of cognitive effort to consciousness is rarely discussed in any detail. One thing to ask, then, is whether cognitive effort is *invariably* conscious.

It is plausible to think that, since many psychological processes (e.g., vision, audition, attention) have conscious and unconscious varietals, cognitive effort may at times be unconscious – cognitive effort may function outside of the agent's awareness. Further elucidation of this idea raises interesting issues. So, for example, Westbrook and Braver (2015) write that 'Cost signals need not always be conscious to influence behavior, but they may become conscious when the costs are sufficiently high' (400). Questions for further research, then, are whether this is true, and if so, what kind of threshold for awareness of cost signals is operative for cognitive control.

Thus far we have discussed the feeling of cognitive effort without considering what the feeling is *like*. Here most assert that cognitive effort feels bad. That is, cognitive effort has an aversive, or negatively valenced, phenomenology. Evidence:

> Why, in other words, don't people always simply perform at the highest level of which they are capable? The intuitive answer suggested by introspection is that we are constitutively reluctant to mobilize all available cognitive resources. That is, mental effort is inherently aversive or costly. (Shenhav et al. 2017, 101)

> [E]ffort is a feeling that emerges in awareness during effortful tasks and reflects the costs associated with goal-directed behavior. (Andre et al. 2019, 1)

> Effortful actions are often attributed to those instances when behavioral control is deployed, are most often thought to be inherently costly in decision-making, be aversive and invoke an urge to disengage even when such actions may be considered adaptive. (Dunn et al 2019, 1033)

> Effort has a distinct phenomenology, feeling difficult and aversive. (Inzlicht et al. 2018, 337)

But even here disagreement is live, with some proposing that cognitive effort may at times be negative and at times positive (Székeley and Michael 2021, Carruthers and Williams 2022). This might lead us to ask what the science of cognitive control indicates about the phenomenology of cognitive effort in particular.[3]

It is difficult to settle on a clear answer here, for the science of cognitive control has not settled on a standard way to measure the phenomenology of cognitive effort. The literature contains a wide range of different strategies, including retrospective questionnaires that probe trait-level features (Venebles and Fairclough 2009, Muraven et al. 2008), designs that explicitly assume a connection between effort and task demand (Botvinick et al. 2009), and self-reports regarding subjective difficulty (Naccache et al. 2005), desire to avoid a task-type (McGuire and Botvinick 2010), amount of effort invested (Mulert et al. 2005), or perceptions of fatigue (see Hagger et al. 2010). It is unclear how exactly some of these measures relate to the phenomenology of cognitive effort, and it is unclear how exactly these different measurements relate to each other.[4]

6 Disagreement and explication

We have covered disagreement and open questions regarding the nature of cognitive effort, regarding the characterization of cognitive effort as an action, and regarding the relationship between cognitive effort and conscious experience. One result of this is that, at present, we have a few options regarding the kind of construct cognitive effort is supposed to be. If experimental papers or experimental designs regarding cognitive effort leave these options unaddressed, there is a risk of cross-talk between researchers, generation of merely verbal disputes, and of accidental changes of subject across different experimental paradigms.

What, then, are we to do? One option is to do nothing. It is possible to argue, for example, that vagueness or a plurality of potential referents is the norm for many of the concepts at use in

---

[3] In this connection, it bears mention that the experiences associated with the 'flow state' are often thought to include neutrally valenced experiences alongside positively valenced experiences (see Shepherd 2021c), and that the flow state is sometimes said to involve an absence of felt effort. One question, then, is whether and how agents achieve flow during the exertion of cognitive control (Huskey et al. 2018).

[4] In this connection, it is worth noting that work a nearby construct, listening effort – basically, cognitive effort directed towards listening tasks – has revealed weak convergent validity between different forms of self-report, physiological, and behavioral measurements (see Johnson et al. 2015, Strand et al. 2021).

cognitive neuroscience (e.g., attention (Anderson 2021), or cognition (Allen 2017)), and that progress in science will naturally clean this up along the way. I accept that it is possible and even fruitful for research on mental capacities to proceed in the face of disagreement about the nature of these capacities. But this *laissez-faire* response is no universal salve. Details matter, and there will be cases where vagueness and underlying disagreements hamper scientific progress (on this point, see Strand et al. 2020).

This might be the case when, for example, experimental design regarding cognitive effort jointly influences things like [a] conscious experience, [b] participant beliefs about effortfulness or difficulty on a task, and [c] task difficulty in a more objective sense. If we are unclear whether research on cognitive effort is studying (e.g.) conscious experience or task difficulty, we might also be unclear whether an experiment warrants an inference about conscious experience or task difficulty. This can easily lead to explanatory bleed, where evidence regarding task difficulty is taken to be evidence regarding conscious experience, and theories and further experiments are constructed that use different measures to probe into the matter, leading to review papers that confidently cite the broad evidential base underlying a claim that was confused from the beginning.

The way forward that I favor is simply to be as explicit and clear as possible regarding the construct guiding experimentation and theory in any given case. So, when discussing 'cognitive effort,' theorists should attempt to go beyond folk psychological characterizations, and should instead specify the features of the construct that are at issue given an experimental set-up or model.[5] Something like this is what Westbrook and Braver (2015) do, in laying out a cognitive effort discounting paradigm. They quantify cognitive effort as 'subjective value,' and specify that subjective value is to be revealed in experiments that measure participant willingness to engage in more or less demanding tasks for more or less reward. They do not, to be clear, explicitly conceive of themselves as offering an explication of 'cognitive effort.' Their view is that free-choice paradigms allow experimenters to measure 'psychophysical intensity' insofar as 'participants, rather than the task parameters, determine responses' (400), and they suggest that this method may be more reliable than methods that rely on participant reports of effort.

This turns attention away from conscious experience, of course. But their clarity is helpful, in part because this clarity gives one a sense of what they are and are not trying to explain. Arguably, it would be possible to utilize experiments in their paradigm alongside experiments that explore the phenomenology of cognitive effort, in order to inform broader questions about cognitive control and consciousness.

On this explicative method, one must take greater care with theoretical terms, and the devil will often be in the details. But if done right, this added clarity might allow one to fruitfully sever reliance on folk psychology, or on semantic intuitions about what qualifies as cognitive effort. This may allow experimentation, model building, and theory construction to proceed without sparking verbal disputes about the 'real nature' of cognitive effort, recognizing instead that many

[5] There is a literature in the philosophy of science regarding 'explication' that addresses best practices in this regard (see Carnap 1950, Justus 2012). An explication is, in a sense, the construction of a technical term suited to specific purposes.

things that have been called cognitive effort may differentially contribute to cognitive control, or to resource allocation.

Consider, by way of illustration, the disagreement canvassed in section 2.1. There are important differences between allocations of amounts of information-processing capacity, selections of levels of intensity for control signals, and optimization-aimed choices of information-processing adjustment. But I submit that there is no great reason to think that any one of these operations uniquely deserves the title cognitive effort. One might instead adopt pluralism about cognitive effort, and maintain that cognitive effort may manifest in more than one way.

Here, then, is a relatively theory-neutral explication of cognitive effort that draws on the above discussion.

Cognitive effort [explication]. Cognitive effort is the intentional or non-intentional deployment and/or maintenance of operations the function of which is to promote greater success at paradigmatic cognitive control tasks.

This explication is pluralistic because it allows that cognitive effort may manifest in a range of ways – a range of control signals related to attention, memory, or transitions in thought (cf. Sripada 2021) may constitute cognitive effort in a given context. This explication is also relatively theory-neutral in that it leaves open questions about what cognitive effort represents, cognitive effort's relationship to action, cognitive effort's relationship to consciousness, and so on. Solid evidence indicating the closure of any of the open questions I have isolated above would allow greater specificity regarding the nature of cognitive effort (see box 1 for a summary of open questions for further research).

The aim of such an explication is, at this juncture, to promote avoidance of verbal disputes about the 'real nature' of cognitive effort, in order to proceed to genuinely open questions regarding the representations and operations involved in cognitive control, the relationship of these representations and operations to conscious experience, to action and automatization, and so on.

Of course, in approaching these open questions we will begin to see subtle interactions between experimental paradigms and background theoretical commitments. Theories of the nature and function of cognitive effort must fit into a bigger picture regarding how cognitive control works, and how to conceive of its constitutive mechanisms. Settling this picture requires more than greater clarity regarding the terms in play – it requires grounding in evidence, and in attempts to coherently integrate evidence drawn from multiple research streams.

In this connection, in this paper's concluding section I wish to emphasize a positive trend in recent cognitive control research. As we will see, this trend is relevant to our understanding of the nature and function of cognitive effort.

7 Conclusion: A broader space of mental value

The positive trend I have in mind involves the recognition of greater complexity and flexibility in the computations and operations performed by our cognitive control capacities. It involves, as well, a placing of cognitive effort within a more complex space of mental value that is sensitive to a number of parameters.

Consider, for example, a recent proposal by Wojtowicz and Loewenstein (2020). They focus on curiosity (see also Chater and Loewenstein (2016)), and argue that hedonic mental states like curiosity function by, in effect, highlighting for decision-making the attractiveness of various candidate mental operations, and doing so by way of cost/benefit computations. As they note, then, 'curiosity operates in a conceptually similar fashion to mental effort and other hedonic states that direct our cognitive strategies' (138).

Evidence is accumulating that both negative and positive conscious feelings play important roles in the guidance of cognitive control. On the positive side, this is demonstrably true of curiosity (Toussaert 2018), confidence (Lee and Daunizeau 2021), and fluency (Oppenheimer 2008). On the negative side, this is true of cognitive effort, and also of boredom (Eastwood et al. 2012), fatigue (Lorist et al. 2005, Agrawal et al. 2021), and anxiety (Eysenck et al. 2007). One important question moving forward, then, is what these experiences have to do with each other. How does the space of mental value, conceived as the range of positive and negative experiences connected with the exercise of cognitive control, influence the operation of cognitive control? And how do individual experiences (e.g., cognitive effort) fit into this broader space?

In this connection, consider a study by Saunders et al. (2015). Experimenters had participants see one of two letters – an M appeared 80 percent of the time (and its presence constituted a low-conflict condition), and a W appeared 20 percent of the time (with its presence constituting a high-conflict condition). They were to press the / key if they saw and M, and the Z key if they saw W. In one condition, after errors, participants received 1s of an aversively high-pitched tone. In a different (unpunished) condition, participants received a quieter and (relatively) non-aversive beep after errors. Participants would undergo 70 trials in a block, and after this they would offer self-reports of their phenomenology along five dimensions. As Saunders et al. explain,

> Three specific questions asked participants to report their affective experience: Anxiety ("How ANXIOUS were you?"), frustration ("How FRUSTRATED were you?"), and hopelessness ("How HOPELESS did you feel?"), while two further questions probed other aspects of phenomenological experience during performance: boredom ("How BORED were you?") and effort ("How HARD did you try?"). (1207)

Their results are interesting. The punishment condition produced higher reports of anxiety, frustration, hopelessness, and effort, as well as lower reports of boredom. In the high-conflict condition (i.e., when participants saw the rarely-presented W), higher participant levels of frustration correlated with higher error rates, while higher participant levels of anxiety correlated with lower error rates (but this latter result was statistically dependent upon the punishment condition). Additionally, when analyzed at the block level (over 70 trials), higher reports of effort were correlated with lower error rates, while higher reports of hopelessness were correlated with higher error rates.

The fact that participants were able to distinguish between all of these elements, many of them plausibly related to conscious experience, coupled with the fact that different experiential elements correlated with different behavioral patterns[6] (different response times and error rates) suggests that this kind of work might prove fruitful for a deeper investigation of the role of different kinds of conscious experiences in the exercise of cognitive control (see Gieseler et al. 2020).

Analyzing the joint role of a range of experiences for cognitive control raises difficult but important questions. How is the space of mental value generated and maintained? One interesting way into this question might look at relationships between metacognitive computations and various evaluative experiences. As already discussed, Carruthers and Williams (2022) argue that cognitive effort stems from model-free metacognitive mechanisms. They also argue that signals of uncertainty are generated by model-free mechanisms. Independently of whether this is true, an interesting question going forward concerns the computational basis of feelings like effort, uncertainty, boredom, curiosity, etc. Do some of these feelings share a computational basis? Are some generated independently by model-free processes? And is it possible for agents to develop metacognitive models that afford decisions based upon comparisons between these various feelings?

Further questions are related to the causes and functions of these feelings. How, for example, do objectively measurable features of cognitive control contexts and performance relate to ongoing conscious experience? What is the connection between, on the one hand, error rates, difficulty, demand, monitoring of context, and monitoring of goal progress (Frömer and Shenhav 2021), and individual differences like need-for-cognition (Inzlicht et al. 2018), level of motivation (Yee and Braver 2018), and self-motivation (Kazén et al. 2015), and, on the other hand, the experiences associated with mental value? Are features that go beyond cost/benefit computations – features like the type of control signals deployed, the learning of context-types, and metacognitive monitoring of aspects like information quality – reflected in consciousness as well? Designing experiments which can illuminate answers to these questions is an important task for those interested in the role of consciousness in cognitive control.

This paper articulates a number of open questions regarding cognitive effort, along several dimensions.

Regarding how effort works. Is the deployment of cognitive effort sometimes ineffective? If so, why? Are there better and worse ways to deploy cognitive effort?

Regarding what effort represents. Are cognitive effort representations model-based or model-free? How closely do representations of cognitive effort track potential causes of cognitive effort – items like opportunity cost, executive engagement, and contextual factors that carry information about these?

Regarding effort's relationship to action. How are we to distinguish between the intentional (or controlled) deployment of cognitive effort, and the non-intentional? To what extent might the selection of cognitive control operations, including decisions about effort investment, proceed automatically? How might agents learn to combine and sequence basic mental actions to achieve sophisticated goals? What kinds of mental flexibility are available to human agents regarding this kind of learning process? What kinds of control signals

[6] In this connection, see Milyavskaya et al. (2019), which found that boredom was closely associated with greater reward sensitivity, while effort was not.

are available for the guidance of attention, memory, or thought? Are these signals are relatively fixed, or is it possible to learn new ways to guide thought over time?

Regarding effort's relationship to consciousness. how are actions of effort investment related to feelings of effort? Is the feeling of effort ever illusory, and why? Is effort ever unconscious, and if so, why do feelings of effort become conscious?

Regarding pluralism about cognitive effort. How many ways might cognitive effort manifest? What are the dimensions – e.g., amount, intensity, control signal type – along which effort may vary, and what is the mechanistic basis for variations along these dimensions?

Regarding the relationship between effort and other evaluative experiences. How is the space of mental value generated and maintained? Is it possible for agents to build (metacognitive) mental models that allow for interactions between, and choices based upon, multiple cognitive feelings? How similar is the computational basis for distinct cognitive feelings (e.g., curiosity, boredom, and cognitive effort)? Are all of these feelings, for example, generated by model-free mechanisms, or do they depend upon model-based computations?

Box 1. Open questions about cognitive effort.

Acknowledgements

My thanks to Wayne Christensen, Carlota Serrahima, Chiara Brozzo, and Wayne Wu for discussion.

Funding information

The author gratefully acknowledges research support from the European Research Council, Starting Grant 757698, awarded under the Horizon 2020 Programme for Research and Innovation. The author also thanks the Canadian Institute for Advanced Research for support in the form of an Azrieli Global Scholar fellowship, and a Catalyst grant on the Functions of Consciousness.

Conflicts of interest

None.

References

Agrawal, M., Mattar, M. G., Cohen, J. D., & Daw, N. D. (2021). The temporal dynamics of opportunity costs: A normative account of cognitive fatigue and boredom. *Psychological review* 129(3), 564-585.

Allen, C. (2017). On (not) defining cognition. *Synthese*, *194*(11), 4233-4249.

Anderson, B. (2021). Stop paying attention to "attention". *Wiley Interdisciplinary Reviews: Cognitive Science*, e1574.

André, N., Audiffren, M. and Baumeister, R.F., 2019. An integrative model of effortful control. *Frontiers in Systems Neuroscience*, *13*, 79.

Badre, D., & Nee, D. E. (2018). Frontal cortex and the hierarchical control of behavior. *Trends in cognitive sciences*, *22*(2), 170-188.

Bayer, U. C., Achtziger, A., Gollwitzer, P. M., & Moskowitz, G. B. (2009). Responding to subliminal cues: do if-then plans facilitate action preparation and initiation without conscious intent?. *Social Cognition*, *27*(2), 183-201.

Botvinick, M. and Braver, T., 2015. Motivation and cognitive control: from behavior to neural mechanism. *Annual review of psychology*, *66*.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S. and Cohen, J.D., 2001. Conflict monitoring and cognitive control. *Psychological review*, *108*(3), 624.

Botvinick, M.M., Huffstetler, S. and McGuire, J.T., 2009. Effort discounting in human nucleus accumbens. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(1), 16-27.

Braver, T. S., & Cohen, J. D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. *Control of cognitive processes: Attention and performance XVIII*, 713-737.

Carnap, R. (1950). *Logical foundations of probability*. Chicago: University of Chicago Press.

Carruther, P., and Williams, D.M. (2022). Model-free metacognition. *Cognition* 225, 105-117.

Chater, N., & Loewenstein, G. (2016). The under-appreciated drive for sense-making. *Journal of Economic Behavior & Organization*, *126*, 137-154.

Chatham, C. H., & Badre, D. (2015). Multiple gates on working memory. *Current opinion in behavioral sciences*, *1*, 23-31.

Chiu, Y. C., & Egner, T. (2017). Cueing cognitive flexibility: Item-specific learning of switch readiness. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(12), 1950.

Churchland, P. M. (1981). Eliminative materialism and propositional attitudes. *the Journal of Philosophy*, *78*(2), 67-90.

Cushman, F., & Morris, A. (2015). Habitual control of goal selection in humans. *Proceedings of the National Academy of Sciences*, *112*(45), 13817-13822.

Dayan, P. (2012). How to set the switches on this thing. *Current opinion in neurobiology*, *22*(6), 1068-1074.

Desender, K., Van Opstal, F. and Van den Bussche, E., 2017. Subjective experience of difficulty depends on multiple cues. *Scientific reports*, *7*(1), 1-14.

Dunn, T.L., Inzlicht, M. and Risko, E.F., 2019. Anticipating cognitive effort: roles of perceived error-likelihood and time demands. *Psychological research*, *83*(5), 1033-1056.

Eastwood, J. D., Frischen, A., Fenske, M. J., & Smilek, D. (2012). The unengaged mind: Defining boredom in terms of attention. *Perspectives on Psychological Science*, *7*(5), 482-495.

Eysenck, M.W., Derakshan, N., Santos, R. and Calvo, M.G., 2007. Anxiety and cognitive performance: attentional control theory. *Emotion*, *7*(2), 336.

Freund, M. C., Etzel, J. A., & Braver, T. S. (2021). Neural coding of cognitive control: The representational similarity analysis approach. *Trends in Cognitive Sciences*, *25*(7), 622-638.

Frömer, R., & Shenhav, A. (2021). Filling the gaps: Cognitive control as a critical lens for understanding mechanisms of value-based decision-making. *Neuroscience & Biobehavioral Reviews*.

Gieseler, K., Inzlicht, M., & Friese, M. (2020). Do people avoid mental effort after facing a highly demanding task? *Journal of Experimental Social Psychology*, *90*, 104008.

Gratton, G., Cooper, P., Fabiani, M., Carter, C. S., & Karayanidis, F. (2018). Dynamics of cognitive control: Theoretical bases, paradigms, and a view for the future. *Psychophysiology*, *55*(3), e13016.

Hagger, M.S., Wood, C., Stiff, C. and Chatzisarantis, N.L., 2010. Ego depletion and the strength model of self-control: a meta-analysis. *Psychological bulletin*, *136*(4), 495.

Herrmann, B., & Johnsrude, I. S. (2020). A model of listening engagement (MoLE). *Hearing Research*, *397*, 108016.

Huskey, R., Craighead, B., Miller, M. B., & Weber, R. (2018). Does intrinsic reward motivate cognitive control? A naturalistic-fMRI study based on the synchronization theory of flow. *Cognitive, Affective, & Behavioral Neuroscience*, *18*(5), 902-924.

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in cognitive sciences*, *22*(4), 337-349.

Inzlicht, M., Schmeichel, B.J. and Macrae, C.N., 2014. Why self-control seems (but may not be) limited. *Trends in cognitive sciences*, *18*(3), 127-133.

James, W. (1890). *The principles of psychology*. London: Macmillan.

Justus, J. (2012). Carnap on concept determination: Methodology for philosophy of science. *European Journal for Philosophy of Science*, *2*(2), 161-179.

Kazén, M., Kuhl, J. and Leicht, E.M., 2015. When the going gets tough…: Self-motivation is associated with invigoration and fun. *Psychological research*, *79*(6), 1064-1076.

Khachouf, O. T., Chen, G., Duzzi, D., Porro, C. A., & Pagnoni, G. (2017). Voluntary modulation of mental effort investment: an fMRI study. *Scientific Reports*, *7*(1), 1-18.

Kool, W., Shenhav, A., & Botvinick, M. M. (2017). Cognitive Control as Cost-Benefit Decision Making. *The Wiley Handbook of Cognitive Control*, 167-189.

Kool, W., & Botvinick, M. (2014). A labor/leisure tradeoff in cognitive control. *Journal of Experimental Psychology: General*, *143*(1).

Kool, W. and Botvinick, M., 2018. Mental labour. *Nature human behaviour*, *2*(12), 899-908.

Kriegel, U. 2015. *The Varieties of Consciousness*. Oxford University Press.

Kurzban, R., 2016. The sense of effort. *Current Opinion in Psychology*, *7*, 67-70.

Kurzban, R., Duckworth, A., Kable, J.W. and Myers, J., 2013. An opportunity cost model of subjective effort and task performance. *Behavioral and brain sciences*, *36*(6), 661-679.

Lee, D. G., & Daunizeau, J. (2021). Trading mental effort for confidence in the metacognitive control of value-based decision-making. *Elife*, *10*, e63282.

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.

Machery, E. (2009). *Doing without concepts*. Oxford University Press.

Massin, O. (2017). Towards a definition of efforts. *Motivation Science*, *3*(3), 230.

McGuire, J. T., & Botvinick, M. M. (2010). Prefrontal cortex, cognitive control, and the registration of decision costs. *Proceedings of the national academy of sciences*, *107*(17), 7922-7926.

Milyavskaya, M., Inzlicht, M., Johnson, T. and Larson, M.J., 2019. Reward sensitivity following boredom and cognitive effort: A high-powered neurophysiological investigation. *Neuropsychologia*, *123*, 159-168.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology*, *41*(1), 49-100.

Mulert, C., Menzinger, E., Leicht, G., Pogarell, O., & Hegerl, U. (2005). Evidence for a close relationship between conscious effort and anterior cingulate cortex activity. *International Journal of Psychophysiology*, *56*(1), 65-80.

Muraven, M., Gagné, M. and Rosman, H., 2008. Helpful self-control: Autonomy support, vitality, and depletion. *Journal of experimental social psychology*, *44*(3), 573-585.

Mylopoulos, M. and Shepherd, J. (2020). The experience of agency. In U. Kriegel (ed.), *Oxford Handbook of the Philosophy of Consciousness*. Oxford University Press.

Naccache, L., Dehaene, S., Cohen, L., Habert, M.O., Guichart-Gomez, E., Galanaud, D. and Willer, J.C., 2005. Effortless control: executive attention and conscious feeling of mental effort are dissociable. *Neuropsychologia*, *43*(9), 1318-1328.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in cognitive sciences*, *12*(6), 237-241.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, 107(1), 179-217.

Pacherie, E., & Mylopoulos, M. (2021). Beyond automaticity: The psychological complexity of skill. *Topoi*, *40*(3), 649-662.

Richter, M., Gendolla, G. H., & Wright, R. A. (2016). Three decades of research on motivational intensity theory: What we have learned about effort and what we still don't know. In *Advances in motivation science* (Vol. 3, pp. 149-186). Elsevier.

Rouault, M. and Koechlin, E., 2018. Prefrontal function and cognitive control: from action to language. *Current opinion in behavioral sciences*, *21*, 106-111.

Saunders, B., Milyavskaya, M. and Inzlicht, M., 2015. What does cognitive control feel like? Effective and ineffective cognitive control is associated with divergent phenomenology. *Psychophysiology*, *52*(9), 1205-1217.

Schmidt, L., Lebreton, M., Cléry-Melin, M. L., Daunizeau, J., & Pessiglione, M. (2012). Neural mechanisms underlying motivation of mental versus physical effort. *PLoS biology*, *10*(2), e1001266.

Shea, N. (2018). *Representation in Cognitive Science*. Oxford University Press.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217-240.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T.L., Cohen, J.D. and Botvinick, M.M., 2017. Toward a rational and mechanistic account of mental effort. *Annual review of neuroscience*, *40*, 99-124.

Shepherd, J. 2016. Conscious action/Zombie action. *Noûs* 50(2), 419-444.

Shepherd, J., 2017. The experience of acting and the structure of consciousness. *The Journal of Philosophy*, *114*(8), 422-448.

Shepherd, J. (2021a). *The Shape of Agency: Control, Action, Skill, Knowledge*. Oxford University Press.

Shepherd, J. (2021b). Skill and sensitivity to reasons. *Review of Philosophy and Psychology*, *12*(3), 669-681.

Shepherd, J. (2021c). Flow and the dynamics of conscious thought. *Phenomenology and the Cognitive Sciences*, 1-20.

Smit, A. S., Eling, P. A., Hopman, M. T., & Coenen, A. M. (2005). Mental and physical effort affect vigilance differently. *International Journal of Psychophysiology*, *57*(3), 211-217.

Sripada, C. (2021). The atoms of self-control. *Noûs*, *55*(4), 800-824.

Strand, J. F., Ray, L., Dillman-Hasso, N. H., Villanueva, J., & Brown, V. A. (2020). Understanding speech amid the jingle and jangle: Recommendations for improving measurement practices in listening effort research. *Auditory perception & cognition*, *3*(4), 169-188.

Székely, M., & Michael, J. (2021). The Sense of Effort: a Cost-Benefit Theory of the Phenomenology of Mental Effort. *Review of Philosophy and Psychology*, 12, 889-904.

Toussaert, S. (2018). Eliciting Temptation and Self-Control Through Menu Choices: A Lab Experiment. *Econometrica*, *86*(3), 859-889.

Venables, L. and Fairclough, S.H., 2009. The influence of performance feedback on goal-setting and mental effort regulation. *Motivation and Emotion*, *33*(1), 63-74.

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 395-415.

Wojtowicz, Z., & Loewenstein, G. (2020). Curiosity and the economics of attention. *Current Opinion in Behavioral Sciences*, *35*, 135-140.

Wu, W. (2013). Mental action and the threat of automaticity. In *Decomposing the Will*, eds. A. Clark, J. Kiverstein, & T. Vierkant, 244-261. Oxford University Press.

Yee, D.M. and Braver, T.S., 2018. Interactions of motivation and cognitive control. *Current opinion in behavioral sciences*, *19*, 83-90.