nature human behaviour

REVIEW ARTICLE
https://doi.org/10.1038/s41562-018-0401-9

Mental labour

Wouter Kool[1] and Matthew Botvinick[2,3]*

Mental effort is an elementary notion in our folk psychology and a familiar fixture in everyday introspective experience. However, as an object of scientific study, mental effort has remained rather elusive. Cognitive psychology has provided some tools for understanding how effort impacts performance, by linking effort with cognitive control function. What has remained less clear are the principles that govern the allocation of mental effort. Under what circumstances do people choose to invest mental effort, and when do they decline to do so? And what regulates the intensity of mental effort when it is applied? In new and promising work, these questions are being approached with the tools of behavioural economics. Though still in its infancy, this economic approach to mental effort research has already uncovered important aspects of effort-based decision-making, and points clearly to future lines of inquiry, including some intriguing opportunities presented by recent artificial intelligence research.

If one listens carefully to the way people talk about mental effort in everyday life, it is striking how often mental effort is tied to terminology from economics. We invest effort, in the hope that the payoff may be worth it. We are exhorted to pay attention. We labour over a decision. Our children work hard at their school work, hopefully not finding it too mentally taxing. Workers at IBM were for many years surrounded by signs reminding them that their job was to THINK!. "We don't get paid for working with our feet," declared their company's founder, "we get paid for working with our heads"[1].

To judge from such locutions, it appears that we associate mental effort with notions of exchange and labour. In short, we think of mental effort in economic terms.

In scientific research on mental effort, there has been a growing trend towards taking this connection seriously, analysing mental effort from an economic point of view and applying formal tools from behavioural economics and labour economics to understand its operation. Even though this approach is relatively new, it has already revealed a number of key features of effort-based decision-making. To underscore this potential, this Review traces the development of this trend, surveys its latest products and considers its future prospects.

The economics of mental effort

In a scientific context, mental effort has been defined as a "subjective intensification of mental activity," which "mediates between how well an organism can potentially perform on some task and how well it actually performs on that task"[2]. From both a psychological and an economic perspective, the question that naturally arises from this definition is, what are the factors that determine the 'intensity of mental activity' that is selected. As we shall review, research pursuing this question has started from two key notions. First, in standard economic fashion, it has been assumed that the allocation of mental effort is rational. That is, people use mental effort to increase payoffs, or in formal economic terms, to maximizing subjective utility. Second, it has been considered that mental effort is associated with negative utility. That is, such effort is costly. Putting these two notions together, the resulting challenge has been to understand mental effort allocation in terms of a cost–benefit analysis, which weighs the potential payoffs of mental effort against its inherent cost.

The economic perspective on mental effort was a natural extension of early behaviourist work, which propounded a law of less work. This stated that "if two or more behavioural sequences, each involving a different amount of energy consumption or work, have been equally well reinforced an equal number of times, the organism will gradually learn to choose the less laborious behaviour sequence leading to the attainment of the reinforcing state of affairs"[3]. Though not phrased in explicitly economic terms, the economic flavour of this proposal is obvious, and subsequent research has cashed out the law of less work in explicit economic terms by mapping it to formalisms from economic labour supply theory (LST)[4].

The law of less work was originally intended to address physical rather than mental effort. However, the notion of effort minimization soon began to pop up in research on cognition and decision-making. In social psychology, Allport[5] explained racial prejudice in these terms, writing, "We like to solve problems easily. Why? Well, it takes less effort, and effort, except in the area of our most intense interests, is disagreeable." Similarly, in discussing the processing of political messages, McGuire[6] characterized human beings as "lazy organisms," seeking to spend as little mental energy as possible. Meanwhile, in cognitive psychology, Baroody and Ginsburg[7] explained strategy choice in arithmetic by appealing to a "drive for cognitive economy." Summing up this point of view, Taylor[8] influentially described human decision-makers as "cognitive misers".

In most early work along these lines, the notion of a cost–benefit trade-off in mental effort was treated informally. However, a pioneering paper[9], 'On the economy of the human-processing system', went further, importing from economics the notion of a utility function, summing the positive utility derived from task performance with the negative utility associated with mental effort itself. Conversely, economics started turning towards psychology, eventually broaching the role of mental effort in decision-making. For example, the economists Smith and Walker[10] introduced a 'labour theory' of economic choice, which featured a term capturing the disutility of cognitive effort, and Camerer and Hogarth[11] put forth a closely related 'capital–labour–production framework', emphasizing the dependence of effort demands on cognitive ability. The latter paper offered a frank expression of the motivating intuition: "Economists instinctively assume thinking is a costly activity. Mental effort is like physical effort — people dislike both."

[1]Department of Psychology, Harvard University, Cambridge, MA, USA. [2]DeepMind, London, UK. [3]University College London, London, UK.
*e-mail: botvinick@google.com

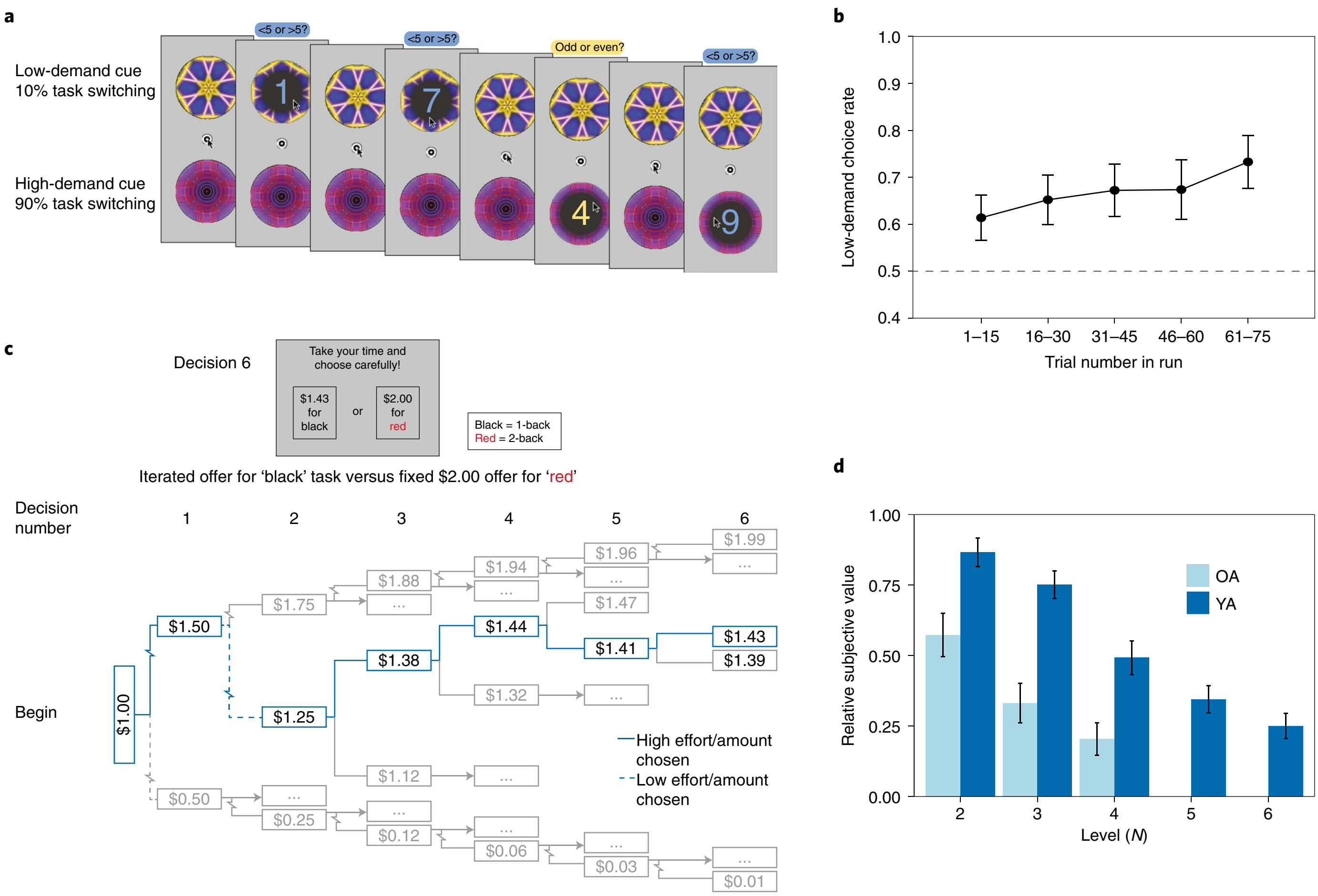


Fig. 1 | Behavioural paradigms that reveal a cost of cognitive control. a, The demand selection task[16]. On each trial, the participant selects between two visual stimuli (left). The selected stimulus reveals a number (right), which is classified by one of two rules depending on its colour. One stimulus (high demand) yields numbers that tend not to repeat the last-encountered colour, requiring frequent effortful task-switching. The other yields numbers that do tend to repeat the last-encountered colour, demanding less effort. **b**, Typical choice behaviour from the demand selection task, showing a tendency to disproportionately choose the low-demand option, as predicted by the labour economic cost-benefit perspective. The error bars indicate s.e.m. and the dashed line indicates chance-level performance. **c**, The cognitive effort discounting (COG-ED) paradigm[17]. Participants chose repeatedly between performing a high-demand task for one amount of dollars and performing a low-demand task for a different amount. The amount offered for the high-demand task was held constant at US$2.00, while the amount offered for the low-demand task was adjusted until participants' choices suggested that they were indifferent between the two offers. **d**, Results from the COG-ED paradigm. Level *N* refers to the level of cognitive demand involved in the high-demand task (an '*N*-back' memory task), with demand increasing to the right. Young adult (YA) participants showed clear 'effort discounting', demanding increasing monetary compensation to sustain increasing mental effort, a measure of the cost associated with cognitive effort. Older adults (OA) also showed effort discounting, but showed greater overall effort costs than the young adults. Error bars indicate s.e.m. Panels adapted from: **b**, ref. [26], National Academy of Sciences; **c**,**d**, ref. [17], PLoS.

Evidence for an intrinsic effort cost

An important limitation attaching to all of the work noted so far is that it leveraged the notion of effort costs as an explanatory principle rather than a hypothesis to be interrogated directly[12–15]. A first step towards overcoming this limitation was made[16], using a task in which participants chose repeatedly between two visual stimuli. When selected, each stimulus produced a series of numbers which, depending on the colour they were presented in, demanded either a magnitude or parity judgement (Fig. 1). The key to the task was that one choice stimulus yielded numbers that required frequent switches between the two classification tasks (high demand), while the other yielded numbers that required less frequent switching (low demand). Across a series of experiments, participants consistently gravitated towards the low-demand option (Fig. 1). This bias was universal across participants, could not be fully explained in terms of error avoidance or minimization of time on task, and was reduced when a monetary incentive was linked with the high-effort choice, confirming its sensitivity to motivational factors.

Closely related work[17–19] provided a similar validation of the notion of cognitive effort costs using a task in which participants choose between performing a low-effort working-memory task for a small monetary reward or a high-effort working-memory task for a larger reward (Fig. 1). After each choice, an adaptive adjustment algorithm changes the offer for the low-effort option until an indifference point between the two options is identified. Across several studies[17–20], this indifference point has been found to link the high-effort option with a higher reward than the low-effort option. The surplus amount associated with the high-reward task can be viewed as quantifying the cost of mental effort.

A number of subsequent studies have adopted similar paradigms to further probe the cost of effort. For example, evidence was provided[21] that cognitive and physical effort are associated with different cost functions. It was shown[22] that effort costs are linked to a

Box 1 | Mental effort and the utility function

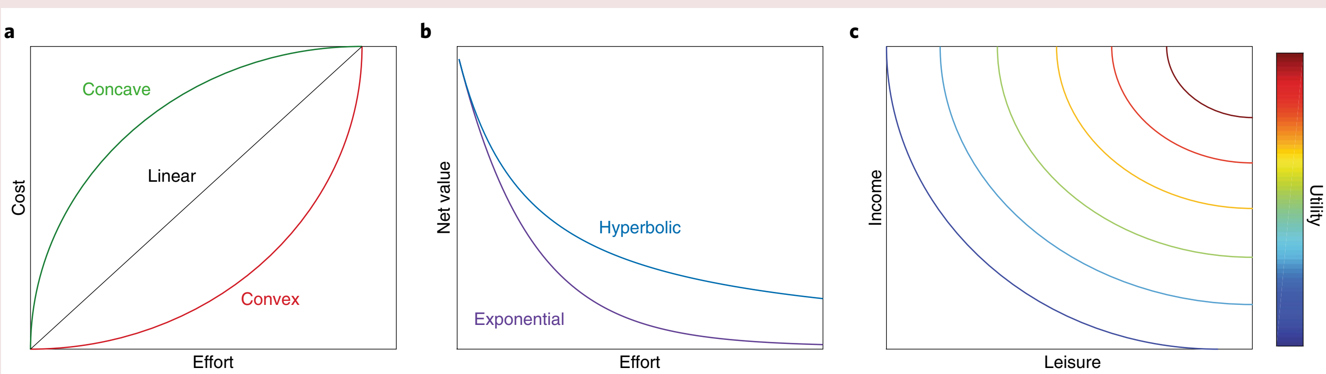
The utility function is a critical construct in economics. Here, a real number (or ordinal rank) is attached to each member of a set of goods or situations, indicating its economic value or 'utility'[40]. The economic approach to cognitive effort has generally modelled the cost of effort by factoring it into the utility function in some manner. In practice, this has taken three broad forms, with further qualitative distinctions within each category.

In the simplest models, it is assumed that the negative utility of effort combines with the positive utility of payoffs in a purely additive manner. This is implicit in models that define a function relating the intensity of effort to a single cost (**a**). In some such models, it is assumed that this relationship is linear (black line in **a**) so that every additional unit of effort adds a fixed amount of disutility to the cost[9,38]. Other work[9] has considered that the form of this function might be convex (red line in **a**), so that successive increments in effort intensity lead to greater increases in the cost of effort. Finally, the relationship between effort and its cost could be concave, so that the cost is relatively sensitive to increases in effort when it is low, but not when it is high. From a behavioural economics standpoint, this form most closely relates to the idea of diminishing returns, with each additional increase in effort leading to smaller increases in its cost.

A second class of models posits more complex, non-additive relationships between rewards and effort costs, but assumes that the impact of effort on reward is nonetheless uniform across reward magnitudes. This form of model takes inspiration from work on temporal discounting — where rewards are decremented by delay — and essentially substitutes effort for time. For example, this relationship could take the form of an exponential discounting function (purple line in **b**), where the effect of additional effort is more comparable across existing levels of effort because of its constant decay rate[95]. Alternatively, it could take the form of a hyperbolic discounting function[21] (blue line in **b**), predicting instead that additional effort more strongly affects the utility of a reward at lower effort demands compared with higher effort demands.

A third class of utility function considers that the effect of effort on utility may depend on both the level of effort and the level of reward, and may assume different functional forms depending on each of these. Inspiration in this case has been drawn from economic LST and, following the convention in that field, effort is usually plotted in terms of its conceptual inverse, 'leisure' and payoffs in terms of cumulative 'income'. Utility then shows up in the third dimension, making the utility function a surface (**c** and see Fig. 2). Note that in the example the iso-utility contours plotted in the panel are concave towards the origin, but that they can take many different forms[40], yielding different behavioural effects, as discussed in the main text.

It is important to realize that current research on effort-based decision-making has not determined which of these functional forms most accurately describes human behaviour. Future lines of research should aim to distinguish between them, and establish functional links between these three different categories of parameterizations of the relationship between effort and value.



Three model classes. a, Several functions linking the intensity of effort to its subjective cost. The green line shows a concave form where each unit of effort incurs less cost, the red line shows a convex form, where cost increases with effort, and the black line shows a linear form where there is a fixed cost per unit of effort. In all cases, effort costs are simply added to rewards to compute net value or utility. **b**, Effort-discounting functions mapping effort to net value. The blue line shows a hyperbolic discounting function, where increases in effort more strongly affect rewards when effort is low compared with when it is high. The purple line shows an exponential discounting function, where increases in effort affect reward similarly across baseline effort levels. **c**, A joint reward–effort utility function, mapping levels of effort (or leisure) and reward (here income) to a net value (plotted in the *z* axis, as indicated by line colours). Here, the iso-utility contours are drawn concave, but these can also take on other forms. Note that the colour bar on the right refers only to **c.**

metacognitive monitoring function, rather than to inherent task demands. These and other studies have helped to clarify which particular kinds of mental activity are associated with effort costs. While effort avoidance has been observed across a fairly wide range of task manipulations[16,17,23–26], the common ingredient appears to be the engagement of top-down cognitive control[27–30], possibly linked with engagement of the brain's executive or 'multiple demand' network[31–34]. This is not surprising, given the longstanding idea that subjective mental effort is linked with the mobilization of cognitive control functions[35,36].

The shape of the effort–cost function

With the notion of a cost of effort now directly validated, several studies have aimed to characterize the 'shape' of the underlying utility function, asking how the subjective cost of effort scales as the intensity of effort itself increases, much as previous work in behavioural economics had asked how the subjective utility of monetary rewards scales with increasing monetary value[37]. While a simple linear model of the cost function has proven adequate to model some behavioural phenomena[9,38], other work points to a nonlinear function. For example, it has been found[24] that people are risk-averse to effort, preferring predictable amounts of effort over risky prospects. Applying a standard behavioural-economic analysis to this finding[37] suggests that the cost function for effort should be convex, with successive increments in effort carrying larger and larger costs. Seeming to contradict this conclusion, a recent study[21] compared several candidate functions and found that the data were best described by a concave (hyperbolic) cost function (Box 1). Further research will be needed to resolve this apparent discrepancy — and also perhaps to explain how hyperbolic discounting might account for decisions to forego both effort and payoff entirely, given that such a discount can never completely erase the value of a prospective payoff. In the meantime, the prevailing uncertainty provides an opportunity to note some inherent difficulties associated with inferring the cost function for mental effort from choice behaviour alone. Crucially, doing so requires the possibility of quantifying effort in terms of some fixed unit. This is inherently difficult, since effort is not a property of the target task alone, but also a function of the individual's cognitive capacities, as well as the degree of effort voluntarily mobilized for the task, which in turn is a function of the individual's reward sensitivity. These factors can be very difficult to disentangle through behavioural observations, a problem which might aptly be called the econometric problem in mental effort research[33,39].

More coherent results have been obtained not by investigating the cost of effort per se, but rather the overall shape of the net utility function that summarizes the effort–reward trade-off. Economic LST has been used to evaluate this[23]. This branch of economics addresses the question of how workers allocate their time between work and leisure, and in particular how this is influenced by hourly wages[4]. The founding assumption of LST is that workers weigh the value of monetary income against the 'value of leisure,' essentially the inverse of the cost of work or effort. Any particular combination of income and leisure time translates to a single utility, and the space of all income–leisure pairings is spanned by a utility 'surface,' as illustrated in Fig. 2. A central tenet of LST is that this utility surface is concave, curved towards the origin and billowing upward as shown in the figure. Although this may sound merely like a technical point, the concavity of the utility function has important consequences for choice behaviour. Most fundamentally, it means that a worker, given an available wage and the choice of how many hours to work, will always favour a balance between income and leisure time (Fig. 2). At a finer grain, the concave shape affects the way that work hours shift in response to increases or decreases in the available wage. In particular, it means that such shifts will depend critically on how many hours are already dedicated to work versus leisure at baseline, and how much income is being already being received[4]. Importantly, non-concave utility functions predict very different effects[4,40]. In particular, a convex or planar utility function would cause workers to always allocate their time entirely to either work or leisure, rather than seeking a mixture of the two that shifts smoothly as wages change; and still other (non-smooth) utility surfaces, studied in economics under the heading of 'perfect complementarity'[40], predict complete insensitivity to wage changes.

Evidence has been provided[23] that mental-effort-based decision-making may cohere with the laws of LST. The concave shape of the LST utility function strongly predicts that a particular form of wage increase, termed an income-compensated increase, will trigger increases in work, while an income-compensated wage reduction will lead to reductions in work (Fig. 2). What makes these predictions non-trivial is that, given the detailed structure of the wage changes, it is possible for decision-makers to work at the same level after the wage change as they did before and attain exactly the same level of income[4] (Fig. 2). The prediction from LST stipulates that workers will forgo this opportunity. This prediction has been successfully tested in studies of labour markets[41–43], providing empirical validation of LST, as well as in animal conditioning experiments looking at physical effort allocation[4]. The same experimental approach was applied to mental-effort-based decision-making[23]. Participants were offered a 'wage' (denominated in pieces of candy) for every minute spent on a cognitively demanding task during a 30 minute test period, with the remainder of the period spent on an unremunerated low-demand task. Income-compensated wage changes induced shifts in effort allocation precisely as predicted by LST (Fig. 2).

These results suggest that the central features of LST, designed originally to account for economic labour markets, may also apply to mental effort. Specifically, they imply that people weigh the income derived from mental effort against the value of mental leisure in a nonlinear fashion, resulting in a preference for balanced combinations of the two. At a finer grain, the results mean that the marginal cost of mental effort is radically baseline-dependent: the cost of effort relative to reward changes depending on how much effort is already being invested and how much income is already being received.

Rewards of mental labour

In other research, economic tools have been leveraged to investigate a range of further issues relating to mental effort. For example, the effort-discounting paradigm has been ported to rodents[44–46], demonstrating a trade-off between cognitive effort and reward in rats. Interestingly, rats show stable individual differences in effort sensitivity[45]. This has also been observed in humans, with effort avoidance and incentive effects on effort correlating with some other cognitive and personality traits[17,47,48] (Fig. 3). Group-level differences have been reported as well, based both on age[17] and psychiatric status[20,49] (Figs. 1 and 3).

One intriguing issue that arises from this individual-differences perspective is that mental effort can, alongside its inherent costs, also carry rewards[2]. For example, in the case of 'learned industriousness'[50], people are proposed to learn a value for the exertion of effort, through years of living in a world where effort often yields rewards. Moreover, the subjective value of earned rewards can sometimes increase when they are obtained through effort[51,52], perhaps reflecting a drive to resolve cognitive dissonance by justifying the investment of effort[53]. An outstanding question is whether these effects reflect an endowment of effort with intrinsic positive value, which effectively cancels some of the inherent cost of the effort, or whether they simply reflect value that is attached to the extrinsic products of the effort. Resolving this question deserves to be a central objective of future research, and the economic framework appears to offer a useful set of tools for accomplishing this.

Strategy selection

One particularly interesting recent trend in research on effort-based decision-making has involved a focus on strategy selection. It has been a long-standing proposal that people select cognitive strategies through a cost–benefit analysis that takes account of effort costs[7,8,54]. Recent work has made progress towards verifying this hypothesis by leveraging computationally explicit models of competing strategies[31,55,56]. Particularly relevant to the present review are cases where strategy choice trades off against economic incentives.

A vivid example of this is provided in a recent study[57]. The focus in this work was on the decision between two strategies for reward-based decision-making, one based on habit and the other based on a more effortful search through possible outcome scenarios[58–61]. In reinforcement learning theory, this distinction has been mapped to a 'model-free' system that simply reinforces actions that have led to reward, and a 'model-based' system that plans over an internal model of the environment towards goals[62–64]. A key question is how decision-makers choose between these two choice mechanisms. The economic perspective on cognitive effort suggests that this arbitration may be the result of a cost–benefit decision, weighing the increased accuracy of model-based planning against its effort costs.

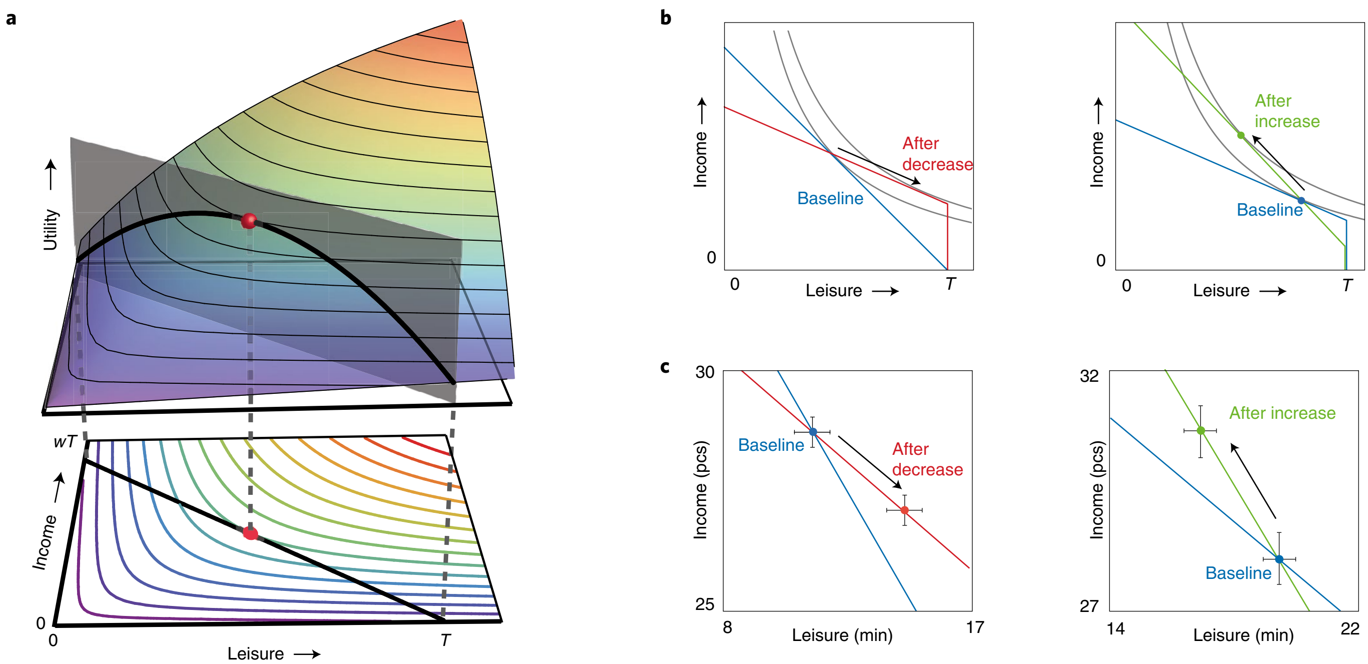


Fig. 2 | Labour supply theory and mental effort. a, Top: a concave utility surface attaching a numerical value to every pairing of income with leisure time. Shown in black are iso-utility or indifference contours, marking out income–leisure pairings with identical utilities. Bottom: a contour map, projected from the indifference curves above. The diagonal line marks out a 'budget constraint,' covering the range of income–leisure combinations open to a worker facing *T* available hours and wage *w*. The grey rectangle above projects this constraint line back up through the three-dimensional utility surface, where the intersection forms an arc. In both top and bottom plots, the red marker highlights the income–leisure combination with the highest utility in the feasible set. **b**, Left: an income-compensated wage reduction. The initial, baseline wage, together with the total available time *T*, defines the initial budget constraint line (blue). The worker chooses the point along this line that maximizes utility (blue point, with associated indifference curve shown in grey). Wage reduction results in a more horizontal budget constraint (diagonal red line). However, it comes along with 'compensating' unearned income (vertical red line), which causes the new budget constraint to intersect the income–leisure combination originally chosen (blue point). Although this combination is thus still available after the wage change, note that the worker also has access to a higher utility combination, which pairs increased leisure with reduced income (red point, with associated iso-utility curve). Right: an income-compensated wage increase, shown in the same fashion. Here, an initial unearned income (vertical blue line) plus an initial wage together define a budget constraint (diagonal blue line). When the wage is raised, the unearned income is reduced (vertical green line), such that the originally chosen income–leisure combination (blue point) is still available. However, the worker now has access to a higher utility combination involving less leisure and greater income (green point). **c**, Experimental results[23] formatted to parallel the plots in **b**. Left: impact of an income-compensated wage reduction on time dedicated to a cognitively easy task (cognitive 'leisure') as opposed to a cognitively demanding task that yielded a wage, denominated in pieces of candy per trial. Right: impact of an income-compensated wage increase. Figure adapted from ref. [23], American Psychological Association.

The study providing the first evidence of this[57] relied on a variant of the so-called two-step task, a behavioural paradigm designed to disentangle model-free from model-based choice (Fig. 4a)[64]. Versions of this task have been used to show that model-based control decreases under cognitive load[65], is related to cognitive control capacity[66] and is reliant on the functioning of prefrontal cortex[67], hallmarks of the kind of computationally expensive mental labour that has been identified as costly. The study[57] tested the cost–benefit model by predicting that behaviour on this would show more model-based control when higher rewards were at stake (Fig. 4b). This pattern was indeed clearly evident in the observed behaviour (Fig. 4c). As a control, the experiment was repeated, but this time using a version of the task in which the application of model-based control does not yield a benefit in terms of rewards[68]. In this setting, the stakes manipulation did not yield a difference in behaviour (Fig. 4c), consistent with the interpretation that decision-makers engage in effortful model-based control only when the prospective payoffs offset the attendant effort costs. In a set of complementary studies[69], the other side of this trade-off was investigated by manipulating the effort costs of model-based control, through the required depth of planning, while keeping the prospective payoffs constant (Fig. 4d–f). Consistent with the cost–benefit account, they found that model-based control was reduced in the face of a deeper, more complex, causal structure. Taken together, these results are consistent with the idea that arbitration between these reinforcement-learning strategies is guided by an economic cost–benefit analysis.

Alongside behavioural studies like these, a growing number of experiments with both humans and animals have applied neuroeconomic techniques to elucidate the neural underpinnings of effort-based decision-making. Although such neuroscientific research falls outside the scope of this Review, it is worth noting that the idea of a cost–benefit analysis has become increasingly central in such work, being formalized, for example, in the proposal that prefrontal circuits maximize the 'expected value of control,' a quantity that takes effort costs into account[31,32]. This and other neuroscientific work on mental effort has been surveyed in a number of recent publications[21,26,31,33,34,70–72].

Why is mental labour costly?

As the existence of a cost of mental effort has become increasingly well validated, and its properties increasingly probed, attention has been increasingly paid to an overarching question: why should mental effort be costly? On first glance this cost seems perverse, since — at least according to our folk psychological notions — effort typically yields positive outcomes. The resulting riddle has not been fully solved. However, here again, an economic perspective has been

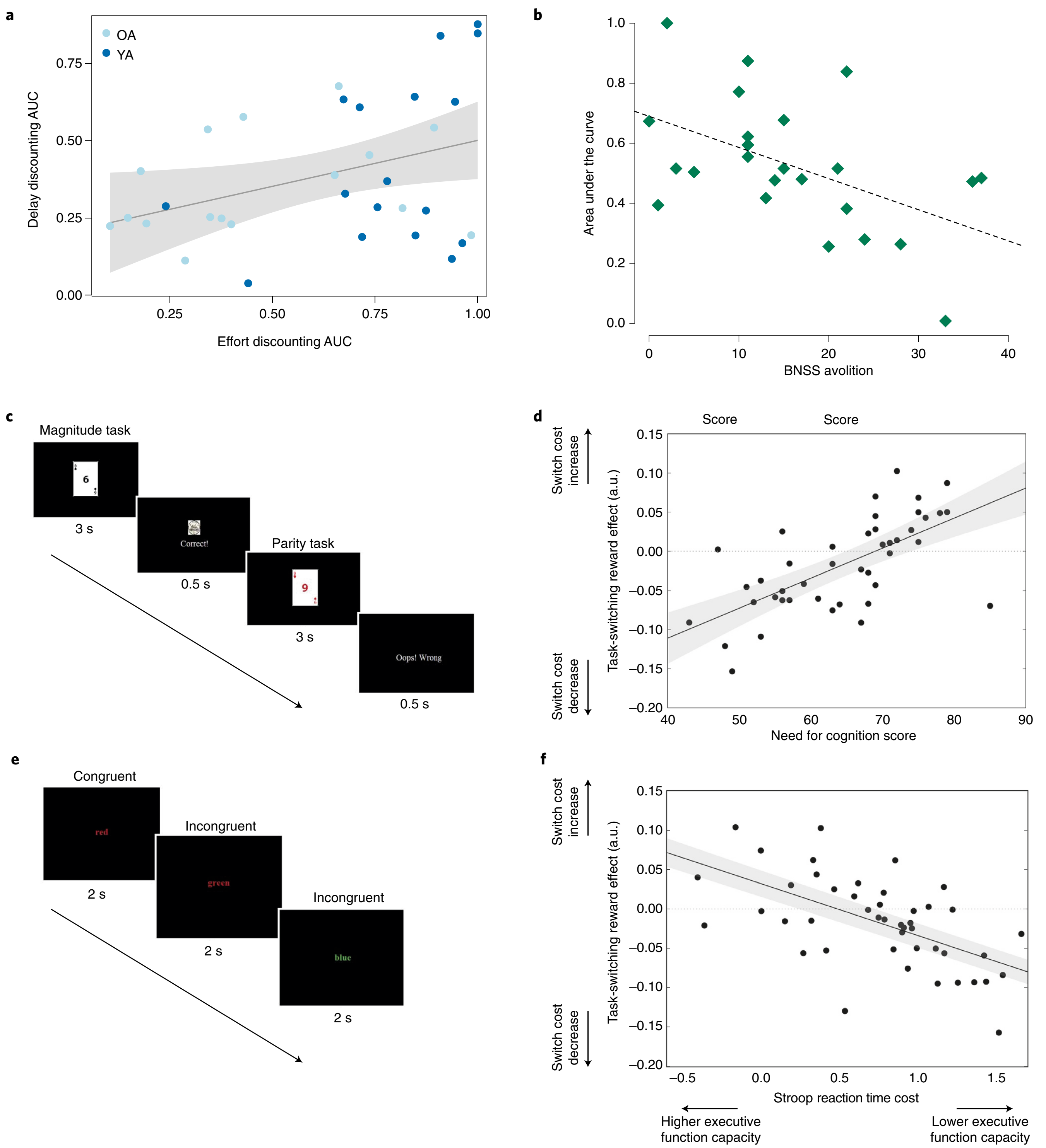


Fig. 3 | Evidence for individual differences correlations in effort-based decision-making. a, A positive correlation was found[17] between the degree of temporal discounting and effort and mental-effort costs as assessed in the COG-ED paradigm (with effort discounting quantified using an 'area under the curve' measurement). This correlation indicates that people with more impulsivity in an intertemporal choice paradigm showed increased mental effort costs on the COG-ED paradigm. The solid line indicates the linear relationship and the shaded area indicates the 95% confidence interval. **b**, Behavioural results[20] revealed a relationship between negative symptoms in schizophrenia (quantified by the brief negative symptom scale (BNSS) avolition scale) and the COG-ED area under the curve measurement of effort costs. This result suggests that motivational deficits typically observed in schizophrenia are related to increased effort discounting. The dashed line indicates the linear relationship. **c–f**, Effort operationalized as the speed with which participants were able to switch between two cognitive tasks[47]. To assess participants' cost–benefit trade-off, they manipulated the monetary incentive of accurate performance from block to block (**c**). They found that increased incentives enhanced effort mainly for participants who scored relatively low on a 'need for cognition' scale, which measures the inclination to engage in demanding cognitive activity in everyday life (**d**). They also found that participants low in cognitive control capacity as measured by the Stroop paradigm (**e**) displayed a larger incentive effect (**f**), presumably because for them larger rewards are needed to offset their increased effort costs. In **d** and **f**, the solid lines indicate the linear relationship and the shaded areas indicate the standard error. Panels adapted from: **a**, ref. [17], PLoS; **c–f**, ref. [47], Elsevier.

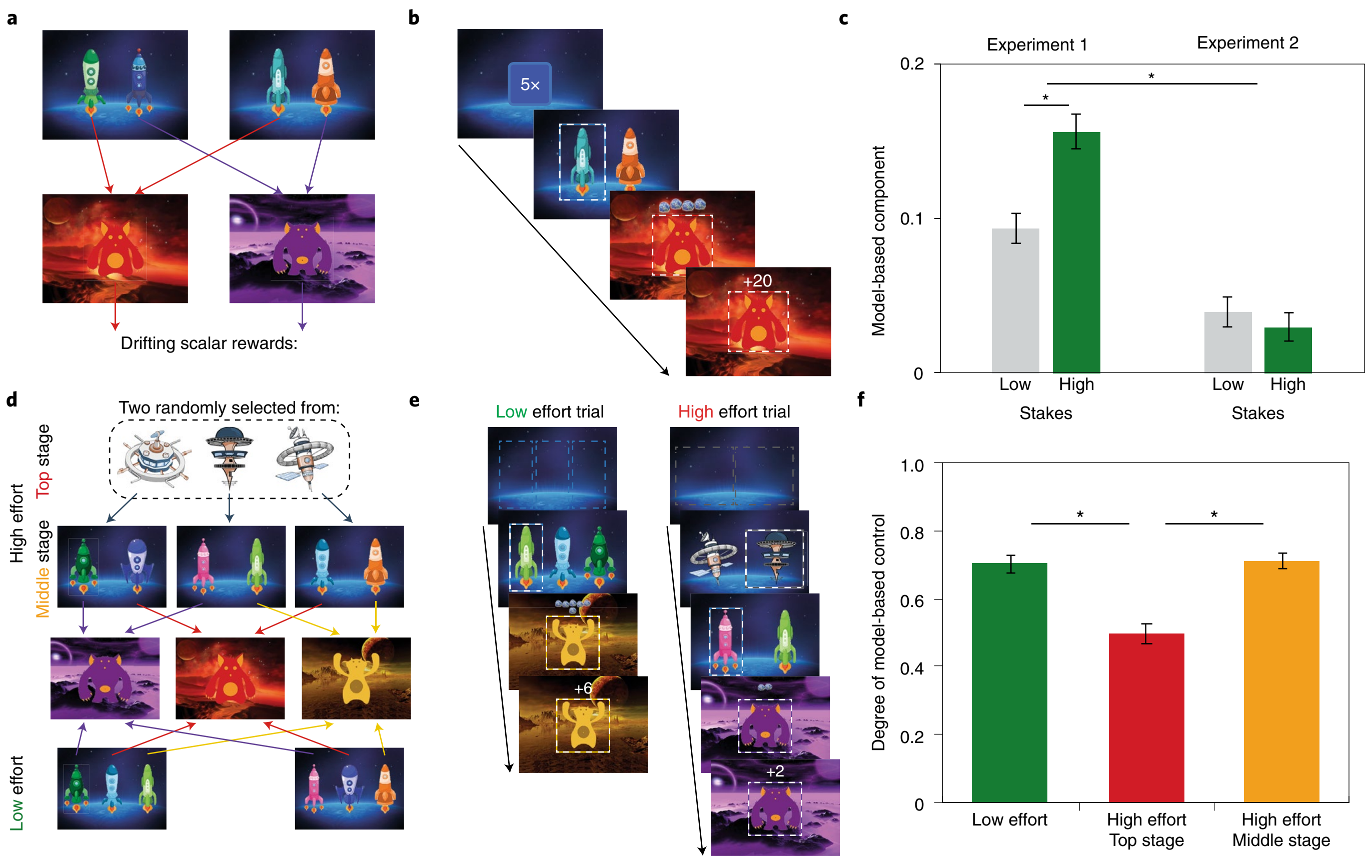


Fig. 4 | Transition structures, trial event sequences and results for multi-stage decision-making tasks. a, Two-step task[68]. Each round of the task begins in one of two first-stage states, each of which offers a choice between two brightly coloured spaceships. Choice of a spaceship triggers a transition — as shown in the figure — to two second-stage states, identified by different aliens. The second-stage states each yield a monetary reward with independent probabilities that change randomly but slowly across trials (to encourage continuous learning). This task distinguishes model-free from model-based control by exploiting the model-based system's knowledge about the equivalence of first-stage states. Model-based decision-makers are able to transfer experiences from one first-stage state to the other. Model-free learners cannot do this, since by definition they rely solely on directly experienced action-reward associations. For example, after a surprisingly large reward on the previous trial, a pure model-based learner will be more inclined to revisit the second-stage state even if the first-stage state in the current trial is not the same as that of the previous trial. A model-free learner is unable to make this inference, because it does not assess actions in terms of their consequences. **b**, To test the effect of increased incentives on model-based control, a stakes manipulation was implemented[57] in the two-step task in **a**. Each trial began with a cue that indicated whether the points on that trial would be multiplied by one (low stakes) or five (high stakes). On the high-stakes trial shown here, the total points earned were 20 (5×4 points). **c**, Model-based choice component for the low- and high-stakes conditions in experiments 1 and 2 from ref. [57]. The 'choice component' measures the influence of model-based decision-making, quantifying this as the increase in probability of choosing the same goal as on the previous trial after positive prediction errors than after negative prediction errors. Prediction errors were estimated using a reinforcement learning model. In experiment 1, where model-based control yielded higher reward, increased stakes triggered stronger reliance on model-based control. In experiment 2, the task was designed such that model-based control yielded no monetary benefit, and in this setting no effect of stakes was observed. **d**, To test the effect of planning demands on model-based control, a novel multi-stage decision-making task was developed[69]. At the start of each trial, the effort condition was randomly selected. On low-effort trials, participants made a single choice between three spaceships that then transitioned to one of three planets. Note that these trials were analogous to the two-step task depicted in **a**, with the exception of an additional planet. High-effort trials first required a choice between two randomly selected space-stations that then transitioned to a pair of spaceships. These spaceships then transitioned to the same final-stage states as in the low-effort trials. **e**, At the start of each trial, empty containers indicate whether the following trial would be either a low- (three containers) or a high-effort (two containers) trial. **f**, A reduction in model-based control in the face of a more complex planning structure was observed, comparing the start of high-effort trials to the start of low-effort trials or the 'middle' stage of high-effort trials[69]. Error bars indicate ±within-subjects s.e.m. and asterisks indicate statistically significant differences derived from *t*-tests (* $P < 0.001$). Panels adapted from: **a**–**c**, ref. [57], SAGE; **d**–**f**, ref. [69], MIT Press.

applied. The general approach is to consider that effort may draw, in some sense, on a scarce resource. Costs, then, can be understood as serving to protect or conserve this resource.

An early hypothesis was that, like physical effort, mental effort might raise metabolic energy consumption. This idea was motivated by a behavioural 'depletion' effect, in which effort was withdrawn after bouts of effort exertion. According to an influential theory, this effect was proposed to reflect a depletion of blood glucose[73]. However, the glucose hypothesis has not held up[74], and the existence of behavioural depletion has itself been called into question[75,76]. There do appear to exist real time-on-task effects on effort[77,78], which interact with incentives[79]. However, these appear to be better explained by motivational variables than by metabolic depletion[80].

A more tenable idea is that mental effort costs are linked to a limited resource that is cognitive or computational in nature, rather

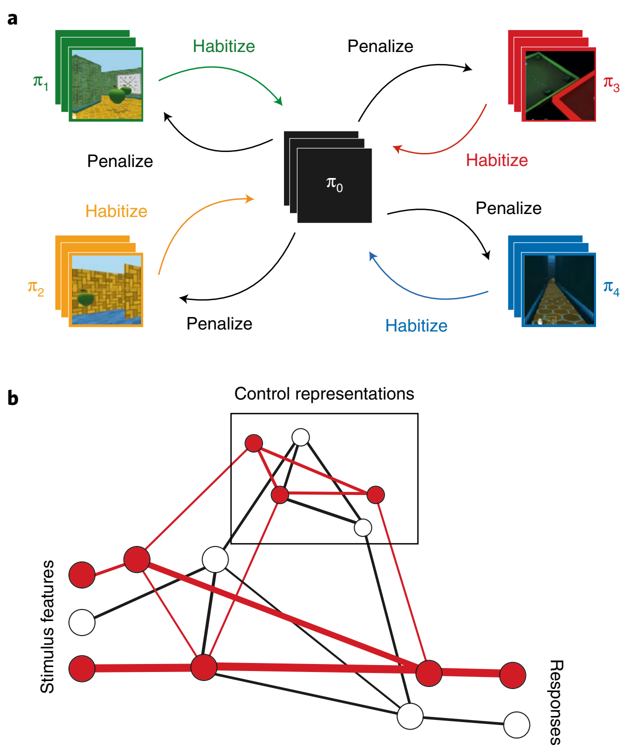


Fig. 5 | Schematic of the proposed multi-task learning model. a, The system[91], which takes the form of a multi-part neural network (not shown), learns several 'policies' or ways of responding to perceptual inputs. These include a 'default' policy (symbolized by π_0) and a set of task-specific policies (π_{1-4}). The system is trained on a set of different tasks, in an interleaved fashion. During this process, the default policy is updated so as to capture elements that the task-specific policies have in common. In the terminology of machine learning, this is a 'distillation' of the task-specific policies into the default policy. In more cognitive terms, it is akin to 'habitizing' the most-common features of the task-specific policies. Meanwhile, the task-specific policies are shaped by a penalty that is imposed when they recommend actions that differ from the actions recommended by π_0. In intuitive terms, the learning process biases the system to 'rely on its habits' as much as possible, consistent with the psychological notion of 'control costs.' **b**, There is a direct and intriguing parallel with psychological models of cognitive control[94]. This features a set of connections between perceptual inputs and action outputs. On their own, these define a set of automatic or habitual behaviours. However, the system also includes 'control' representations, which bias or modulate processing in the habit pathway, allowing habits to be overridden in favour of non-routine responses. By analogy to the proposed architecture[91], the habit pathway implements the default policy, while control inputs give rise to task-specific policies. The penalty term[91] translates, in this context, into a cost of top-down cognitive control. Panels adapted from: **a**, ref. [91], MIT Press, **b**, ref. [94], Annual Reviews.

than metabolic[9]. In some work, this resource has been characterized broadly in terms of computation or information processing[54,55], and indeed human decision-makers do display a tendency to select information processing strategies or algorithms that minimize computational demands[55]. Computation inherently demands time, and another related perspective on effort costs is that they are tied to time through 'opportunity costs,' the rewards associated with foregone activities[81–83]. Another form of opportunity cost derives from the exploration–exploitation dilemma, the inherent trade-off between exploiting established knowledge and exploring the environment so as to gain new knowledge[84]. It has been proposed[80] that effort avoidance may reflect an adaptive bias that triggers disengagement from exploitative cognitive tasks to encourage exploration (see also ref. [30]).

Another, not incompatible, perspective ties the limited cognitive resource more specifically to cognitive control. Control functions have long been understood as a bottleneck in cognitive information processing[55,85,86], providing a potential explanation for why subjective costs appear to be specifically linked with their allocation. Indeed, a number of cognitive modelling enterprises have explicitly incorporated a principle of 'minimal control'[87,88] or 'least-effort,' again referring to cognitive control[54,89].

Having covered these points, it is worth noting that, on close inspection, the idea of a limited resource is not quite sufficient to explain the existence of mental effort costs. In principle, a rational allocation strategy should be able to make the most of such a limited resource (whatever its character), without the need for an inherent cost. To address this issue, it was suggested[90] that the cost of mental effort might serve as a heuristic, wired in by evolution, which discourages the wasteful allocation of cognitive control without requiring detailed allocation decisions that might themselves be computationally intensive (see also ref. [81]).

Another, more novel possibility can be advanced based on recent artificial intelligence research leveraging deep reinforcement learning. Multi-task learning in a neural network architecture that learned both a generic, default stimulus–response policy and a set of task-specific policies has been investigated[91] (Fig. 5). Crucially, during learning, the default policy was updated by 'distilling' the collection of task-specific policies, absorbing their common features into the default policy, while the task-specific policies were shaped by a cost term that penalized them for departing from the default policy. The rationale for this setup was that it would encourage the emergence of a maximally useful default policy, and simulation results showed that the learning architecture displayed greater stability and reliability than architectures without the relevant features.

Although this work was not addressed to human psychology, it gives rise to an intriguing new idea concerning the cost of mental effort. By analogy to standard cognitive models of automatic versus controlled behaviour[29,31], the default policy in the recently reported architecture[91] can be thought of as implementing a set of automatic or habitual behaviours, and the task-specific policies a set of controlled behaviours, which override the automatic behaviours when appropriate (Fig. 5). From this point of view, the simulation results[91] suggest that it may be beneficial to attach a cost to top-down cognitive control specifically because it encourages the development of optimal habits, resulting in a system that relies on control as little as possible and that displays reliable, stable learning in multi-task environments (see also ref. [92]).

Concluding remarks

The economic perspective on mental effort has had a decidedly beneficial impact on psychological research. It has stimulated the development of important new tools, which are now being exploited to shed new light on many aspects of mental-effort-based decision-making. At the same time, it has revealed novel questions that might not otherwise have been considered and has suggested new theoretical perspectives. It seems likely that the economic approach to mental effort will remain productive, continuing to provide leverage on important issues. At the same time, recent work suggests that the economic approach may benefit from ideas stemming from other fields, and in particular ideas from reinforcement learning theory and optimal control. In this regard, the recent explosion in research on artificial intelligence[93] may provide a useful font of ideas.

Received: 8 January 2018; Accepted: 18 July 2018;
Published online: 3 September 2018

References

1. Belden, T. G. & Belden, M. R. *The Lengthening Shadow: The Life of Thomas J. Watson* (Little, Brown, Boston, 1962).
2. Inzlicht, M., Shenhav, A. & Olivola, C. The effort paradox: effort is both costly and valued. *Trends Cogn. Sci.* **22**, 337–349 (2017).
3. Hull, C. L. *Principles of Behavior* (Appleton-Century, New York, NY, 1943).
4. Kagel, J. H., Battalio, R. C. & Green, L. *Economic Choice Theory: An Experimental Analysis of Animal Behavior* (Cambridge Univ. Press, Cambridge, 1995).
5. Allport, G. W. *The Nature of Prejudice* (Addison Wesley, New York, NY, 1954).
6. McGuire, W. J. in *The Handbook of Social Psychology* (eds Lindzey, G. & Aronson, E.) 136–314 (Addision-Wesley, Reading, 1969).
7. Baroody, A. J. & Ginsburg, H. P. in *Conceptual and Procedural Knowledge: The Case of Mathematics* (ed. Hiebert, J.) 75–112 (Lawrence Erlbaum Associates, Hillsdale, 1986).
8. Taylor, S. E. in *Cognition, Social Behavior, and the Environment* (ed. Harvey, J. H.) 189–211 (Erlbaum, Hillsdale, 1981).
9. Navon, D. & Gopher, D. On the economy of the human-processing system. *Psychol. Rev.* **86**, 214–255 (1979).
10. Smith, V. L. & Walker, J. M. Monetary rewards and decision cost in experimental economics. *Econ. Inq.* **31**, 245–261 (1993).
11. Camerer, C. F. & Hogarth, R. M. The effects of financial incentives in experiments: a review and capital-labor-production framework. *J. Risk Uncertainty* **19**, 7–42 (1999).
12. Shah, A. K. & Oppenheimer, D. M. Heuristics made easy: an effort-reduction framework. *Psychol. Bull.* **134**, 207–222 (2008).
13. Shugan, S. M. The cost of thinking. *J. Consum. Res.* **7**, 99–111 (1980).
14. Simon, H. A. A behavioral model of rational choice. *Q. J. Econ.* **69**, 99–118 (1955).
15. Payne, J. W., Bettman, J. R. & Johnson, E. J. Adaptive strategy selection in decision making. *J. Exp. Psychol.* **14**, 534–552 (1988).
16. Kool, W., McGuire, J. T., Rosen, Z. & Botvinick, M. M. Decision making and the avoidance of cognitive demand. *J. Exp. Psychol.* **139**, 665–682 (2010).
17. Westbrook, A., Kester, D. & Braver, T. S. What Is the subjective cost of cognitive effort? Load, trait, and aging effects revealed by economic preference. *PLoS ONE* **8**, e68210 (2013).
18. Westbrook, A. & Braver, T. S. Cognitive effort: a neuroeconomic approach. *Cogn. Affect. Behav. Neurosci.* **15**, 395–415 (2015).
19. Westbrook, J. A. & Braver, T. S. The economics of cognitive effort. *Behav. Brain Sci.* **36**, 704–705 (2013).
20. Culbreth, A., Westbrook, A. & Barch, D. Negative symptoms are associated with an increased subjective cost of cognitive effort. *J. Abnorm. Psychol.* **125**, 528–536 (2016).
21. Chong, T. T.-J. et al. Neurocomputational mechanisms underlying subjective valuation of effort costs. *PLoS Biol.* **15**, e1002598 (2017).
22. Dunn, T. L., Lutes, D. J. & Risko, E. F. Metacognitive evaluation in the avoidance of demand. *J. Exp. Psychol. Hum. Percep. Perform.* **42**, 1372–1387 (2016).
23. Kool, W. & Botvinick, M. A Labor/leisure tradeoff in cognitive control. *J. Exp. Psychol. Gen.* **143**, 131–141 (2014).
24. Apps, M. A., Grima, L. L., Manohar, S. & Husain, M. The role of cognitive effort in subjective reward devaluation and risky decision-making. *Sci. Rep.* **5**, 16880 (2015).
25. Manohar, S. G. et al. Reward pays the cost of noise reduction in motor and cognitive control. *Curr. Biol.* **25**, 1707–1716 (2015).
26. McGuire, J. C. & Botvinick, M. Prefrontal cortex, cognitive control, and the registration of decision costs. *Proc. Natl Acad. Sci. USA* **107**, 7922–7926 (2010).
27. Posner, M. I. & DiGirolamo, G. J. in *The Attentive Brain* (ed. Parasuraman, R.) 401–423 (MIT Press, Cambridge, MA, 1998).
28. Shiffrin, R. M. & Schneider, W. Controlled and automatic information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychol. Rev.* **84**, 127–190 (1977).
29. Botvinick, M. M. & Cohen, J. D. The computational and neural basis of cognitive control: charted territory and new frontiers. *Cogn. Sci.* **38**, 1249–1285 (2014).
30. Schouppe, N., Ridderinkhof, K. R., Verguts, T. & Notebaert, W. Context-specific control and context selection in conflict tasks. *Acta Psychol.* **146**, 63–66 (2014).
31. Shenhav, A., Botvinick, M. M. & Cohen, J. D. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* **79**, 217–240 (2013).
32. Shenhav, A., Cohen, J. D. & Botvinick, M. M. Dorsal anterior cingulate cortex and the value of control. *Nature Neurosci.* **19**, 1286–1291 (2016).
33. Botvinick, M. & Braver, T. Motivation and cognitive control: from behavior to neural mechanism. *Annu. Rev. Psychol.* **66**, 83–113 (2015).
34. Sayalı, C. & Badre, D. Neural systems of cognitive demand avoidance. *Neuropsychologia* https://doi.org/10.1016/j.neuropsychologia.2018.06.016 (2018).
35. Kahneman, D. *Attention and Effort* (Prentice-Hall, Englewood Cliffs, 1973).
36. Posner, M. I. & Snyder, C. R. R. in *Information Processing and Cognition: The Loyola Symposium* (ed. Solso, R. L.) 55–85 (Erlbaum Associates, Hillsdale, 1975).
37. Glimcher, P. W., Camerer, C. F., Fehr, E. & Poldrack, R. A. *Neuroeconomics: Decision Making and the Brain* (Academic, Oxford, 2009).
38. Verguts, T., Vassena, E. & Silvetti, M. Adaptive effort investment in cognitive and physical tasks: a neurocomputational model. *Front. Behav. Neurosci.* **9**, 57 (2015).
39. Musslick, S., Cohen, J. D. & Shenhav, A. Estimating the costs of cognitive control: theoretical validation and potential pitfalls. In *Proc. 40th Annu. Meeting Cognitive Science Society* (2017).
40. Nicholson, W. & Snyder, C. M. *Microeconomic Theory: Basic Principles and Extension* (Cengage Learning, Mason, OH, 2008).
41. Charness, G. & Kuhn, P. *What Can Labor Economists Learn from the Lab?* NBER Working Paper No. 15913 (2010).
42. Dickinson, D. L. An experimental examination of labor supply and work intensities. *J. Labor Econ.* **17**, 638–670 (1999).
43. Fehr, E., Goette, L. & Zehnder, C. A behavioral account of the labor market: the role of fairness concerns. *Annu. Rev. Econ.* **1**, 355–384 (2009).
44. Hosking, J. G., Cocker, P. J. & Winstanley, C. A. Prefrontal cortical inactivations decrease willingness to expend cognitive effort on a rodent cost/benefit decision-making task. *Cereb. Cortex* **26**, 1529–1538 (2015).
45. Cocker, P. J., Hosking, J. G., Benoit, J. & Winstanley, C. A. Sensitivity to cognitive effort mediates psychostimulant effects on a novel rodent cost/benefit decision-making task. *Neuropsychopharmacology* **37**, 1825–1837 (2012).
46. Hosking, J. G., Floresco, S. B. & Winstanley, C. A. Dopamine antagonism decreases willingness to expend physical, but not cognitive, effort: a comparison of two rodent cost/benefit decision-making tasks. *Neuropsychopharmacology* **40**, 1005–1015 (2015).
47. Sandra, D. A. & Otto, A. R. Cognitive capacity limitations and need for cognition differentially predict reward-induced cognitive effort expenditure. *Cognition* **172**, 101–106 (2018).
48. Chevalier, N. Willing to think hard? The subjective value of cognitive effort in children. *Child Dev.* **89**, 1283–1295 (2018).
49. Culbreth, A. J., Moran, E. K. & Barch, D. M. Effort-based decision-making in schizophrenia. *Curr. Opin. Behav. Sci.* **22**, 1–6 (2018).
50. Eisenberger, R. Learned industriousness. *Psychol. Rev.* **99**, 248–267 (1992).
51. Olivola, C. Y. & Shafir, E. The martyrdom effect: when pain and effort increase prosocial contributions. *J. Behav. Decis. Making* **26**, 91–105 (2013).
52. Norton, M. I., Mochon, D. & Ariely, D. The IKEA effect: when labor leads to love. *J. Consum. Psychol.* **22**, 453–460 (2012).
53. Festinger, L. *A Theory of Cognitive Dissonance* (Stanford Univ. Press, Stanford, CA, 1962).
54. Anderson, J. R. *The Adaptive Character of Thought* (Erlbaum, Hillsdale, 1990).
55. Shenhav, A. et al. Toward a rational and mechanistic account of mental effort. *Ann. Rev. Neurosci.* **40**, 99–124 (2017).
56. Lieder, F. & Griffiths, T. L. Strategy selection as rational metareasoning. *Psychol. Rev.* **124**, 762–794 (2017).
57. Kool, W., Gershman, S. J. & Cushman, F. A. Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychol. Sci.* **28**, 1321–1333 (2017).
58. Kahneman, D. Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* **93**, 1449–1475 (2003).
59. Dickinson, A. Actions and habits: the development of behavioural autonomy. *Phil. Trans. Roy. Soc. Lon. B* **308**, 67–78 (1985).
60. Wason, P. C. & Evans, J. S. B. Dual processes in reasoning? *Cognition* **3**, 141–154 (1975).
61. Sloman, S. A. The empirical case for two systems of reasoning. *Psychol. Bull.* **119**, 3–22 (1996).
62. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).
63. Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80**, 312–325 (2013).
64. Daw, N., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–1215 (2011).
65. Otto, A. R., Gershman, S. J., Markman, A. B. & Daw, N. D. The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol. Sci.* **24**, 751–761 (2013).
66. Otto, A. R., Skatova, A., Madlon-Kay, S. & Daw, N. D. Cognitive control predicts use of model-based reinforcement learning. *J. Cogn. Neurosci.* **27**, 319–333 (2014).
67. Smittenaar, P., FitzGerald, T. H., Romei, V., Wright, N. D. & Dolan, R. J. Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron* **80**, 914–919 (2013).
68. Kool, W., Cushman, F. A. & Gershman, S. J. When does model-based control pay off? *PLoS Comput. Biol.* **12**, e1005090 (2016).

69. Kool, W., Gershman, S. & Cushman, F. A. Planning complexity registers as a cost in metacontrol. *J. Cogn. Neurosci.* https://doi.org/10.1162/jocn_a_01263 (2018).
70. Botvinick, M. M., Huffstetler, S. & McGuire, J. C. Effort discounting in human nucleus accumbens. *Cogn. Affect. Behav. Neurosci.* **9**, 16–27 (2009).
71. Schmidt, L., Lebreton, M., Cléry-Melin, M.-L., Daunizeau, J. & Pessiglione, M. Neural mechanisms underlying motivation of mental versus physical effort. *PLoS Biol.* **10**, e1001266 (2012).
72. Frobose, M. I. et al. Catecholaminergic modulation of the avoidance of cognitive control. Preprint at *bioRxiv* https://doi.org/10.1101/191015 (2017).
73. Gailliot, M. T. et al. Self-control relies on glucose as a limited energy source: willpower is more than a metaphor. *J. Pers. Soc. Psychol.* **92**, 325–336 (2007).
74. Inzlicht, M., Schmeichel, B. J. & Macrae, C. N. Why self-control seems (but may not be) limited. *Trends Cogn. Sci.* **18**, 127–133 (2014).
75. Randles, D., Harlow, I. & Inzlicht, M. A pre-registered naturalistic observation of within domain mental fatigue and domain-general depletion of self-control. *PLoS ONE* **12**, e0182980 (2017).
76. Hagger, M. S. et al. A multilab preregistered replication of the ego-depletion effect. *Perspect. Psychol. Sci.* **11**, 546–573 (2016).
77. Braun, D. A. & Arrington, C. M. Assessing the role of reward in task selection using a reward-based voluntary task switching paradigm. *Psychol. Res.* **82**, 54–64 (2017).
78. Blain, B., Hollard, G. & Pessiglione, M. Neural mechanisms underlying the impact of daylong cognitive work on economic decisions. *Proc. Natl Acad. Sci. USA* **113**, 6967–6972 (2016).
79. Brewer, G. A., Lau, K. K., Wingert, K. M., Ball, B. H. & Blais, C. Examining depletion theories under conditions of within-task transfer. *J. Exp. Psychol.* **146**, 988–1008 (2017).
80. Inzlicht, M. & Schmeichel, B. J. in *The Handbook of Self-Regulation* (eds Vohs, K. D. & Baumeister, R. H.) 165–181 (Guilford, New York, NY, 2016).
81. Kurzban, R. The sense of effort. *Curr. Opin. Psychol.* **7**, 67–70 (2016).
82. Kurzban, R., Duckworth, A., Kable, J. W. & Myers, J. An opportunity cost model of subjective effort and task performance. *Behav. Brain Sci.* **36**, 661–679 (2013).
83. Otto, A. R. & Daw, N. The opportunity cost of time modulates cognitive effort. *Neuropsychologia* https://doi.org/10.1016/j.neuropsychologia.2018.05.006 (2017).
84. Wilson, R. C., Geana, A., White, J. M., Ludvig, E. A. & Cohen, J. D. Humans use directed and random exploration to solve the explore–exploit dilemma. *J. Exp. Psychol.* **143**, 2074–2081 (2014).
85. Feng, S. F., Schwemmer, M., Gershman, S. J. & Cohen, J. D. Multitasking versus multiplexing: Toward a normative account of limitations in the simultaneous execution of control-demanding behaviors. *Cogn. Affect. Behav. Neurosci.* **14**, 129–146 (2014).
86. Musslick, S. et al. Controlled vs. automatic processing: a graph-theoretic approach to the analysis of serial vs. parallel processing in neural network architectures. In *Proc. 38th Annu. Conf. Cognitive Science Society* 1547–1552 (2016).
87. Taatgen, N. A. in *Integrated Models of Cognitive Systems* (ed. Gray, W. D.) 368–379 (Oxford Univ. Press, New York, NY, 2007).
88. Yeung, N. & Monsell, S. Switching between tasks of unequal familiarity: the role of stimulus-attribute and response-set selection. *J. Exp. Psychol. Hum. Percep. Perform.* **29**, 455–469 (2003).
89. Gray, W. D. The nature and processing of errors in interactive behavior. *Cogn. Sci.* **24**, 205–248 (2000).
90. Kool, W. & Botvinick, M. The intrinsic cost of cognitive control. *Behav. Brain Sci.* **36**, 697–698 (2013).
91. Teh, Y. et al. Distral: Robust multitask reinforcement learning. *Adv. Neural Inf. Process. Syst.* **30**, 4497–4507 (2017).
92. Genewein, T., Leibfried, F., Grau-Moya, J. & Braun, D. A. Bounded rationality, abstraction, and hierarchical decision-making: An information-theoretic optimality principle. *Front. Robot. AI* **2**, 27 (2015).
93. Hassabis, D., Kumaran, D., Summerfield, C. & Botvinick, M. Neuroscience-inspired artificial intelligence. *Neuron* **95**, 245–258 (2017).
94. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).
95. Białaszek, W., Marcowski, P. & Ostaszewski, P. Physical and cognitive effort discounting across different reward magnitudes: tests of discounting models. *PLoS ONE* **12**, e0182353 (2017).

Competing interests

The authors declare no competing interests.

Additional information

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence should be addressed to M.B.