Measuring the time utility of mental effort

Samarth Mehrotra (samarthm20@iitk.ac.in)
Department of Cognitive Science,
IIT Kanpur, India

Nisheeth Srivastava (nsrivast@iitk.ac.in)
Department of Computer Sciencee and Engineering,
IIT Kanpur, India

Abstract

The empirical measurement of mental effort is an important problem in the cognitive sciences. Recently, researchers have adopted econometric tools to attempt to characterize mental effort in terms of monetary costs foregone. Such efforts yield a very helpful calculation device - a money utility of mental effort. However, since the opportunity cost of applying mental effort in any given situation is measured with respect to time rather than money in most ecologically reasonable settings, it is even more desirable to obtain a measure of the time utility of mental effort. In the absence of direct measurements of mental effort, such a task has proved econometrically challenging. We use a recently developed direct measure of mental effort to characterize its time utility, finding that it is approximately linear in effort. We discuss some implications of this result for current theories of mental effort, as well as for practical applications.

Keywords: decision making; decisions from experience; optimal information sampling; mental effort; visual estimation; microeconomics

Introduction

The utility function is a fundamental mathematical device in the analysis of micro-economic behavior (Dominick, 2008), with the metaphor of utility extending deep into the study of behavior more generally, with ideas of utility maximization and risk aversion richly informative about peoples' behavior in both natural and artificial settings (Marschak, 1950; Luce & Raiffa, 1989).

It has recently been realized that it is possible to meaningfully characterize the phenomenon of mental effort, previously mostly conceptualized in philosophical (Conlisk, 1996) or biophysical terms (Mulder, 1986; Fairclough & Houston, 2004) in an economic framework resembling utility maximization. We 'invest' our mental effort with things like education and we 'pay' for tasks like taking care of our young ones with mental effort (Kool & Botvinick, 2018), an insight substantiated by empirical results supporting avoidance of cognitive demand (Kool, McGuire, Rosen, & Botvinick, 2010). This line of work has solidified the appropriateness of using the economic metaphor of 'cost' in the study of mental effort - something that we prefer to reduce, given a choice (Navon & Gopher, 1979).

On this account, positive utility accrued from performing a mental task is offset by a *negative utility or cost* incurred by mental effort. By permitting the operationalization of mental effort in the same mental currency units as monetary preferences or other overt markers of value, such accounts enable econometric methods to be used to characterize the arithmetic relationship between these negative costs and positive utilities.

The shape of this curve has interesting theoretical implications. One possibility is a linear relationship where every additional amount of effort adds a fixed cost, implying that people are willing to continue working harder with better economic incentives. Empirically measured labour-leisure indifference curves, however, are convex in shape (Mankiw et al., 1997), where every unit of effort contributes an increasing cost to the equation suggesting there is some 'satisficing' solution for effort level in any given task (Schwartz et al., 2002). A convex relationship where increasing effort has decreasing cost would imply, counterintuitively, that greater effort is actually attractive rather than aversive.

The recently developed cognitive effort discounting (henceforth COG-ED) experimental paradigm uses monetary reward to measure utility of mental effort, which they call 'subjective value' with the understanding, as we discuss above, that mental effort can be measured in the same mental currency as task utility (Westbrook, Kester, & Braver, 2013). In this experimental paradigm, a n-back memory task is used to stimulate varying levels of effort, and the participants is given trade offers between doing a base level task (1-back) at a lower reward or a higher n-back task for a higher reward. Keeping the the higher reward and base level task fixed, the compensation for the lower task is modulated at every individual n-back task (n=2 to 6), until the subject is indifferent between doing the higher and lower n-back task. This obtained indifference value normalised by the higher reward magnitude yields the subjective value, which serves as a proxy for mental effort.

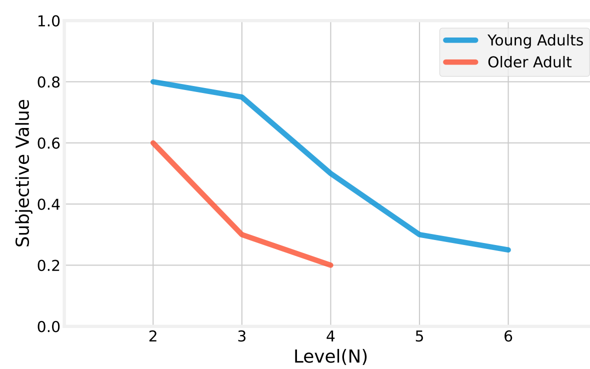

Figure 1: Subjective Value of Old Adults(OA) and Young Adults(YA). Re-plotted from Westbrook et al. (2013)

Figure 1, re-plotted from Westbrook et al. (2013), shows a monotonically decreasing relationship of subjective value (increasing costs) with increasing mental effort, as assessed using the COG-ED paradigm. However, attempts to construct a relationship curve between these variables requires that the response characteristics of both variables be approximately linear internally. Money-based objective markers typically display concave response characteristics with respect to subjective valuations (Seidl, 2013), while the response characteristics of n-back memory tasks with respect to mental effort are unknown since we have no reliable direct measurement of the latter. Thus, econometric analyses of curves obtained using COG-ED and related methods cannot proceed assuredly.

The growing mental effort literature appreciates this conundrum in the form of an 'econometric problem of mental effort'(Kool & Botvinick, 2018) pointing to two prominent confounds in existing measurements of mental effort: a) individual differences between processing the same information and b) subjective allocation of mental effort; which make its measurement non-trivial.

Solving the econometric problem

In a nutshell, mental effort while making decisions is conceived in information processing terms to operationalize it as an object capable of quantitative study(Shenhav et al., 2017). But how much information is processed while doing a task also depends on the confounds we describe above. Translated to the n-back memory task, participants can choose to process more information by trying to remember details inessential for the completion of task. Even if they do the task the same way, they display individual differences in how efficiently they process information in each step, with practice enabling a reduction of effort for the same level of performance (Shiffrin & Schneider, 1977). Therefore completion of an n-back trial for a fixed n might take different levels of mental effort for different subjects, or even for one subject across different trials.

Further, even if this problem is solved, every additional unit of the independent variable (x-axis in Figure 1) should correspond to the same additional unit of dependent variable (y-axis), if we are concerned with characterising the shape of the curve. In the simplest case, scales should be linear. It is uncertain whether the expected mental effort difference between 1-back and 2 back task is same as that between 2-back and 3-back task; certainly the response dynamics become highly non-linear as participants approach their working memory capacity limits.

Thus, a solution to the econometric problem requires experimenters to use effort tasks with low inter-subject variability, low potential for across-trial learning, and low potential for distraction. To this end, in the experiment we report in this paper, we replaced the n-back task in the COG-ED paradigm with a recently developed visual estimation task (Mehrotra & Srivastava, 2022) that, we demonstrate in our work, provides an approximately linear scale for mental effort.

Additionally, instead of measuring the subjective utility of mental effort in terms of monetary units we measure it with respect to time. The opportunity cost of mental effort is naturally measured in time units, since mental effort precludes doing something else at any point in time. Therefore, measuring the utility of mental effort in time units makes more ecological sense than doing so in monetary units. Thus, in this paper, we present a new approach to measuring the time utility of mental effort.

The visual estimation task

Mehrotra and Srivastava (2022) designed a visual estimation task, in turn derived from visuomotor lotteries designed in Juni, Gureckis, and Maloney (2011), where the participant has to find the location of a hidden circle on the screen with help of hints. The hints come in form a small dots, the position of which is determined through samples from a bi-variate Gaussian distribution with a fixed standard deviation centered on the hidden circle. The dots appear one at a time and remain on the screen until a guess is made. The way to generate these hints was to make a visual fixation of one second on the screen. Each hint also cost a fixed amount which was reduced from the initial reward. The idea of having an external cost was to have both an explicit cost and benefit associated with hints, which would allow mathematical ascertainment of an optimal trade-off between these two, finding the optimal number of hints that this trade-off occurs at and finally comparing the participants behaviour with this value.

Promisingly for our purpose, this task makes considerable progress towards solving both the parts of econometric problem of constructing a linear scale of measurement of mental effort. The action of generating hints is mapped to visual fixations of 1000ms which are found to be insignificantly variant inter-subjectively(Shiffrin & Schneider, 1977). The number of hints sampled provides a simple parametric measure of effort. Imposing explicit costs for each hint, , as in Mehrotra and Srivastava (2022), allows an optimal number of hints to be calculated. With increasing number of hints, the probability of success increased and the potential reward decreases. Expected reward can, in turn, be calculated as a function of number of hints by multiplying these two measures and the optimal number of hints (Figure 2).

For our purposes, we modify this task by removing the cost element from it and impose repetition of the trial until success is achieved, shifting the nature of cost from intangible 'money' to tangible 'time'. With the number of hints sampled remaining the unit of effort, manipulating the expected number of hints for optimal performance would mean manipulating the level of effort, which was the driving force of our study. Measuring trade-able time at the indifference point at multiple such levels of effort maps out a time utility curve for a single participant. To do this, we had to recalculate the expected number of hints in this modified setup.

Calculating optimal effort

The original task set the difficulty or the level of effort required to complete the task by changing the standard deviation of the bi-variate Gaussian distribution from which the location of hints was sampled. This distribution was centered on the hidden circle and thus the point hints that appeared essentially hinted at the probable center of the hidden circle.

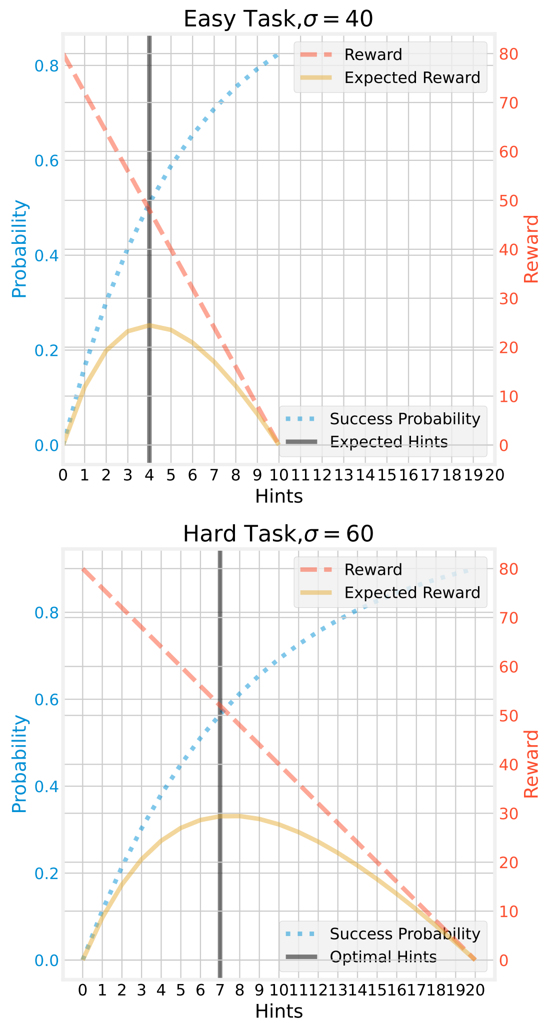

Figure 2: Expected number of hints for tasks of different difficulty levels. The mode of the success probability curve gives us the optimal number of hints for a given difficulty level.

Generating a few such hints that would stay on the screen would give sampling distribution of the center of hidden circle thus indicating a probable area at which the participant could take a guess.

In a given trial of our modified visual estimation task, the participant had to repeat the task until they become successful. The probability of hitting the target in n hints as a function of σ, the standard deviation was:

$$P[hit/n] = P(\sigma) = \int\int_{T} \phi(0, \Sigma) dxdy,$$

where T is the area of the hidden circle and ϕ is the probability density function of the multivariate Gaussian with Σ as covariance matrix:

$$\Sigma(n) = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & \sigma^2/n \end{bmatrix}$$

Participants were free to sample as many hints as they wanted before making a guess, but with errors leading to the process restarting. In this process, the expected number of hints for a given effort level (σ) is simply the probability of being successful at the current attempt and not being successful at any previous one multiplied by the number of hints taken in total, summed over all possible number of attempts and hints. This can be specified by the following equation:

$$EH(\sigma) = \frac{1}{L_n} \sum_{n=1}^{L_n} n \sum_{a=1}^{L_a} aP[hit/n] \prod_{0}^{a-1} (1 - P[hit/n]),$$

where $EH(\sigma)$ is the expected hints which is the function of σ, n is the number of hints, a is the number of attempts, L_n is upper limit of number of hints and L_a as the upper limit of number of attempts. An analytical solution of this equation above would consider the summations over infinity, instead of some arbitrary limits, but an appropriate numerical approximation estimated from observational data serves our purpose, as we show further below. Modulating the expected hint count as a function of effort level construct a scale for mental effort, whose properties we also examine in more detail further below.

COG-ED with visual estimation

We incorporated the visual estimation task designed above in the cognitive effort discounting paradigm (Westbrook et al., 2013) which in our case ascertains an indifference point where the participant would be indifferent between doing the visual estimation task at a particular effort level or simply waiting for a stated duration of time. For a given effort level, a participant was given an option to do the task or wait (wait task) for time t which was sampled from the range 0-40 seconds in factors of 5. In the wait task, participants had to fixate at a fixation cross for the said duration. If participants accepted the time trade, all lower values from the range was discarded from further sampling transitively assuming lower values to also be preferred against a task at the current effort level. Similarly on rejection of wait task upper values were discarded. This processes was repeated until a single value was left in the range and this value was regarded as the point of indifference - a value at which the participant is approximately indifferent to doing the visual estimation task or the wait task.

Methods

Here, we describe methodological details for an experiment we conducted using the COG-ED visual estimation task.

Participants

30 university students (11 females, 19 males; mean age: 22.3) participated in the experiment for monetary compensation (100). All participants reported perfect eyesight. Data was collected with approval from the university's Institutional ethics committee.

Apparatus

The experiment was displayed on a 1920x1080 pixel screen in a dark room. A standard PC mouse was used to click and guess the position of the target. An Eyelink 1000 eye tracker was used to record gaze data at 1000Hz. A head mount was used to fix the position of the head. The PsychoPy python library was used to create the stimuli and Pylink was used to integrate the eye tracker.

Stimuli

For the visual estimation task, a circle of radius 24 pixels was used as the target. Hints were dots of 4 pixels radius. The standard deviation used for five effort levels was 70 to 110 pixels. The standard deviations were not revealed to the participants, only the 'difficulty level'. At the beginning of every trial, two cards (1/8 of the screen size) were displayed in the middle of either side of the screen, displaying two choices: wait 'T' seconds or do the visual estimation task at 'D' difficulty. A correct guess revealed the circle in green, otherwise red, after the participant had responded. In the latter case, we generated the visual estimation task again at the same difficulty level but with a different location of the hidden target. The fixation box inside which the participants needed to foveate for a minimum of 1000 ms was a rectangle with 1/4 the height and width of the screen. For the wait task, the fixation cross was 200 pixels in height and width.

Design

The factor of 'expected hints' with 5 levels indexed the effort-demand level. The dependent variable of 'the point of indifference' was measured in time units. Participants had to do each level once, with the number of fixation tasks and estimation tasks varying within each level, based on participant choices.

Procedure

The participants began with the familiarisation phase where they had to succeed in the visual estimation task at least five times at every effort level. The average time taken to succeed at all levels was equally weighted with each individual level's time to create a reference point for each level. In the main phase, they were shown all the relevant stimuli and instructions and were told what randomly picked 'difficulty level' they would be going through. A fixation box appeared randomly positioned on the screen, and they had to bring their gaze inside it and fixate for 1000 ms to generate a dot like hint. They could generate as many hints as they wanted, and when they were confident, they had to point and click with the mouse where they thought the circle was, and then the circle revealed itself in either green or red colour based on if the guess was correct or not, respectively. If incorrect, the trial started again with the fixation box in a new randomly generated position. They had to repeat this for every difficulty level.

In the main phase, they were given a choice between two options presented on two cards. They could either choose to do nothing for 'T' seconds, in which case they would have to fixate on a cross for that said duration, or they could do the task at 'D' difficulty. For a given difficulty level, this was repeated until the indifference point was found.

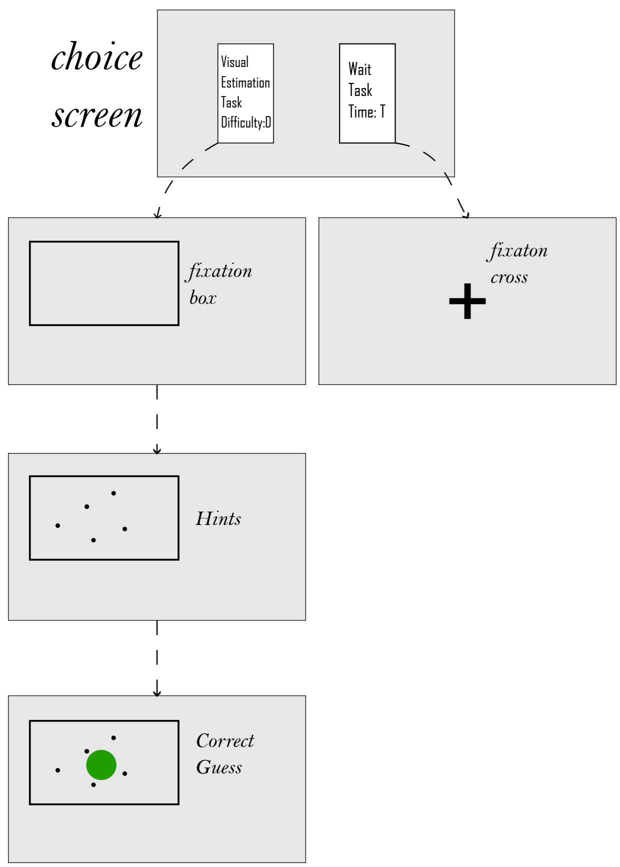

Figure 3: Illustration of the experiment setup. Upon taking a choice the participant either does the visual estimation task or the wait task.

Results

An effort scale with linear response characteristics

Figure 4 shows the expected hints curve for effort levels for the upper limit of possible hints set at L_h at 20,30 and 50 and the upper limit of possible attempts (failures + 1) L_a set at at 10 and 20. For these limits, varying the standard deviation σ of the generative distribution of hints between 70 to 110 in increments of 10 provides an almost linear scale, indicated by R_2 fit values of around 99 when plotting SD vs expected

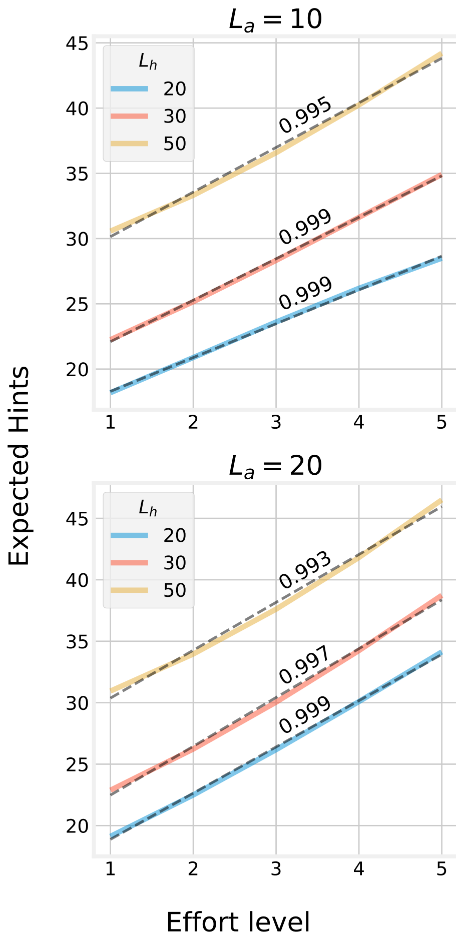

Figure 4: Linear fit across different limits. Text on line shows the R^2 values for the best fit line.

effort. Thus, we obtain 5 values of the independent variable on a linear scale, with the assurance that they vary linearly with the expected effort at each level.

Respondents undersample hints

Figure 5 summarizes participants' sampling behavior vis-a-vis the expected sampling behavior estimated using L_h and L_a limits estimated from the data itself. For the participant data L_h=22 at the 90^{th} percentile across all participants, with a maximum of L_h=27 seen for one participant. We set $L_h = 27$ to err on the side of caution. $L_a = 4$ at the 90^{th} percentile, and is set as such to calculate the expected effort curve. As we see in Figure 5, participants under-sample across effort levels, in contrast to earlier observations of oversampling seen in Juni et al. (2011) and Mehrotra and Srivastava (2022). We reflect upon this incongruity further below.

The time utility curve is approximately linear

The primary goal of this study was to estimate the time utility curve for mental effort. To achieve this, we fit a variety of mathematical functions to the effort level vs number

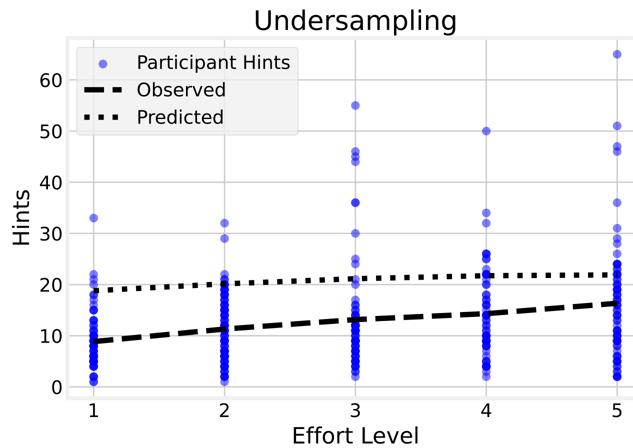

Figure 5: Observed hints curve compared with empirically estimated expected hints curve.

of hints data collected from our participants. The following curve families were fit by minimizing the least square error; best fit curves are shown in Figure 6:

1. Linear: $y = ax + b$
2. Convex: An exponential curve: $y = a(1+b)^{x-c} + d$
3. Concave: A sigmoidal curve: $y = \frac{a}{1+e^{-b(x-c)}} + d$

AICs for model fits are 2800,2800 and 2794 for the convex, concave and linear curves respectively, suggesting that the time utility curve of mental effort, at least on the time and effort scales measured in our experiment, is approximately linear, in accord with similar linear measurements of the money utility curve for mental effort previously documented in Westbrook et al. (2013). R^2 value for each curve was 0.12. Here we are not concerned with predictions but rather the shape of data, thus low R^2 values don't effect our results.

Discussion

In this paper, we presented a method for measuring the time utility of mental effort, repurposing a recently proposed mental effort measurement to trade off effort for time instead of effort for money. An experimental evaluation of this method revealed three observations. One, that the method is econometrically reasonable in the sense that it produces linear response characteristics between task difficulty and expected effort level. Two, that respondents under-sample relative to the expected effort level. Three, that the measured time utility curve is approximately linear.

Our observation of systematic under-sampling is in contrast with (Mehrotra & Srivastava, 2022) and (Juni et al., 2011) who have observed consistent oversampling in the same visual estimation task. It is interesting to note that Juni et al. (2011) describe purpose of their study as an

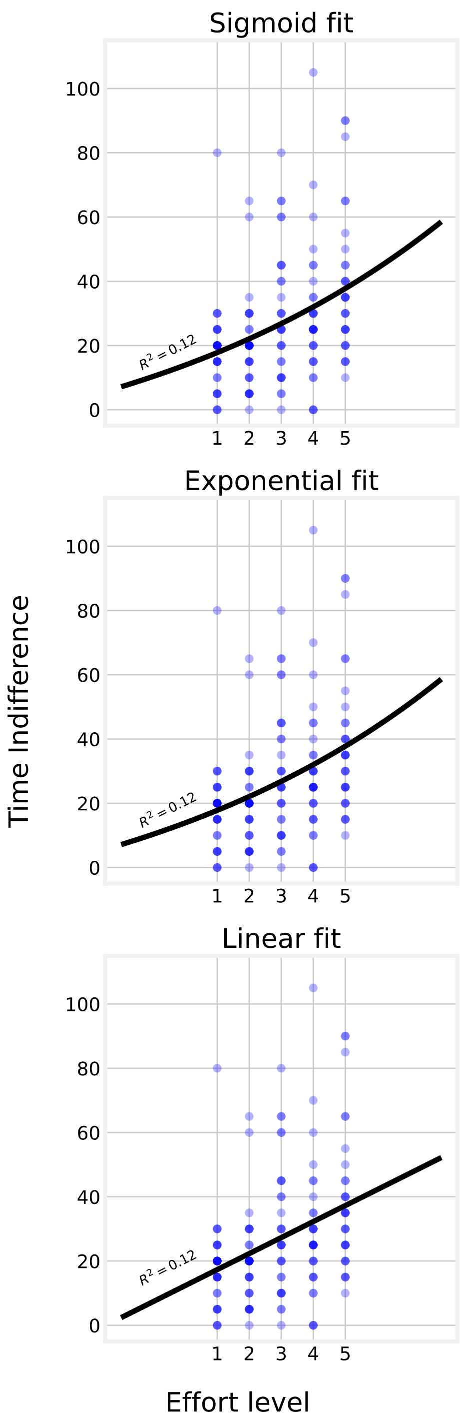

Figure 6: Three curve families fit to the data.

effort to measure optimal information sampling behaviour under explicit costs, as a move away from the then en vogue paradigm of information sampling studies without explicit costs, which have reliably demonstrated undersampling (Hertwig & Pleskac, 2010). We find this trend again in our study in the absence of explicit costs, reinforcing the argument made in Juni et al. (2011). It appears that individuals under sample evidence when costs are imposed in terms of time, and over-sample when an explicit cost structure in terms of quasi-monetary units is presented, consistent with time being treated as a more fungible good than money, but also with many other theoretical possibilities. A deeper investigation of this difference constitutes an interesting avenue for future work.

Our experimental data are consistent with a linear shape for the time utility of mental effort on time-scales and effort-scales consistent with lab experiments (up to one minute of moderately effortful activity). That is, in this setting at least, every increasing unit of mental effort has the same utility for individuals if the utility is measured with time, suggesting that neither respondents' time nor effort budgets are constrained by the task we set them.

It is almost certain that this conclusion cannot generalize to longer time-scales, wherein diminishing returns are certain to arise, or to other task contexts wherein the shape of the curve may be a function of expertise with the task and other factors (Meijman, 1997). Nonetheless, the time and effort scale of our experiment is entirely consistent with lab-scale experimentation, and also seems likely to generalize to important practical use cases, such as the design of computer-human interfaces (Zugal, Pinggera, Reijers, Reichert, & Weber, 2012; Baumeister et al., 2017), wherein measurements of time-effort trade offs are an explicit need, and in pedagogical assessments that relate mental effort with difficulty (Srivastava, Srivastava, & Chandrasekharan, 2020, 2021).

Finally, while mental effort may be operationalized in information processing terms to render it better suited for quantitative and computational reasoning (Shenhav et al., 2017), it is important to remember that this operationalization is limited, and ignores considerable phenomenological and practical detail about the nature of mental effort, e.g. the fact that the same amount of effort can 'feel' starkly different based on motivation factors (Kleiber, Larson, & Csikszentmihalyi, 1986). It remains important, therefore, to caveat the discussion about the measurement of mental effort to the specific aspects of it that are in fact amenable to an information processing metaphor.

References

Baumeister, J., Ssin, S. Y., ElSayed, N. A., Dorrian, J., Webb, D. P., Walsh, J. A., ... others (2017). Cognitive cost of using augmented reality displays. *IEEE transactions on visualization and computer graphics*, *23*(11), 2378–2388.

Conlisk, J. (1996). Why bounded rationality? *Journal of economic literature*, *34*(2), 669–700.

Dominick, S. (2008). *Principles of microeconomics.* Oxford University Press, New Delhi.

Fairclough, S. H., & Houston, K. (2004). A metabolic measure of mental effort. *Biological psychology*, *66*(2), 177–190.

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition*, *115*(2), 225–237.

Juni, M., Gureckis, T., & Maloney, L. (2011). Don't stop 'til you get enough: adaptive information sampling in a visuomotor estimation task. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 33).

Kleiber, D., Larson, R., & Csikszentmihalyi, M. (1986). The experience of leisure in adolescence. *Journal of leisure research*, *18*(3), 169–176.

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

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of experimental psychology: general*, *139*(4), 665.

Luce, R. D., & Raiffa, H. (1989). *Games and decisions: Introduction and critical survey*. Courier Corporation.

Mankiw, N. G., et al. (1997). *Introduction to microeconomics*. Harcourt Brace College Publishers.

Marschak, J. (1950). Rational behavior, uncertain prospects, and measurable utility. *Econometrica: Journal of the Econometric Society*, 111–141.

Mehrotra, S., & Srivastava, N. (2022). Selecting between visuomotor lotteries to measure mental effort in risky decisions. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44).

Meijman, T. F. (1997). Mental fatigue and the efficiency of information processing in relation to work times. *International Journal of Industrial Ergonomics*, *20*(1), 31–38.

Mulder, G. (1986). The concept and measurement of mental effort. *Energetics and human information processing*, 175–198.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological review*, *86*(3), 214.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: happiness is a matter of choice. *Journal of personality and social psychology*, *83*(5), 1178.

Seidl, C. (2013). The st. petersburg paradox at 300. *Journal of Risk and Uncertainty*, *46*, 247–264.

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

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: Ii. perceptual learning, automatic attending and a general theory. *Psychological review*, *84*(2), 127.

Srivastava, A., Srivastava, N., & Chandrasekharan, S. (2020). Measure conceptmapping, not concept-maps: Procedural analysis elucidates stages in students' understanding of biology concepts. In *annual meeting of the american educational research association. retrieved july.*

Srivastava, A., Srivastava, N., & Chandrasekharan, S. (2021). Quality of concept maps is affected by map building strategies. *Biochemistry and Molecular Biology Education*, *49*(1), 129–139.

Westbrook, A., Kester, D., & Braver, T. S. (2013). What is the subjective cost of cognitive effort? load, trait, and aging effects revealed by economic preference. *PloS one*, *8*(7), e68210.

Zugal, S., Pinggera, J., Reijers, H., Reichert, M., & Weber, B. (2012). Making the case for measuring mental effort. In *Proceedings of the second edition of the international workshop on experiences and empirical studies in software modelling* (pp. 1–6).