OPEN MIND
Discoveries in Cognitive Science
an open access journal

Optimal Indolence: A Normative Microscopic Approach to Work and Leisure

Ritwik K. Niyogi[1], **Yannick-Andre Breton**[2], **Rebecca B. Solomon**[2], **Kent Conover**[2], **Peter Shizgal**[2] **and Peter Dayan**[1]

[1]Gatsby Computational Neuroscience Unit, University College London, London, United Kingdom,

[2]Center for Studies in Behavioral Neurobiology, Concordia University, Montreal, Quebec, Canada

Keywords: work, leisure, normative, microscopic, reinforcement learning, economics

ABSTRACT

Dividing limited time between work and leisure when both have their attractions is a common everyday decision. We provide a normative control-theoretic treatment of this decision that bridges economic and psychological accounts. We show how our framework applies to free-operant behavioural experiments in which subjects are required to work (depressing a lever) for sufficient total time (called the price) to receive a reward. When the microscopic benefit-of-leisure increases non-linearly with duration, the model generates behaviour that qualitatively matches various microfeatures of subjects' choices, including the distribution of leisure bout durations as a function of the payoff. We relate our model to traditional accounts by deriving macroscopic, molar, quantities from microscopic choices.

Citation: Niyogi, R. K., Breton, Y.-A., Solomon, R. B., Conover, K., Shizgal, P., Dayan, P. (2015). Optimal indolence: a normative microscopic approach to work and leisure. Open Mind 1(1): 112.

DOI:
http://dx.doi.org/10.1098/rsif.2013.0969

Supplemental Materials:
http://dx.doi.org/10.1098/rsif.2013.0969

Received: 20 October 2013
Accepted: 7 November 2013
Published: 26 January 2014

Competing Interests: The authors have declared that no competing interests exist.

Corresponding Author:
Ritwik K. Niyogi
ritwik.niyogi@gatsby.ucl.ac.uk

The MIT Press

INTRODUCTION

What to do, when to do it and how long to do it for are fundamental questions for behaviour. Different options across these dimensions of choice yield different costs and benefits, making for a rich, complex, optimization problem.

Sample Subsection

One common decision is between working (performing an employer-defined task) and engaging in leisure (activities pursued for oneself). Working leads to external rewards such as food and money; whereas leisure is supposed to be intrinsically beneficial (otherwise one would not want to engage in it). Since these activities are usually mutually exclusive, subjects must decide how to allocate time to each. Note that work need not be physically or cognitively demanding, but consumes time; equally leisure need not be limited to rest, and may present physical and/or mental demands.

Sample Subsubsection Note that work need not be physically or cognitively demanding, but consumes time; equally leisure need not be limited to rest, and may present physical and/or mental demands.

Symbol	Meaning
$1/\lambda$	mean of exponential effective prior probability density for leisure time
$\alpha \in [0,1]$	weight on linear component of microscopic benefit-of-leisure
$\beta \in [0,\infty)$	inverse temperature or degree of stochasticity-determinism parameter
CHT	Cumulative Handling Time
$C_L(\cdot)$	microscopic benefit-of-leisure
$C_{L_{max}}$	maximum of sigmoidal microscopic benefit-of-leisure
$C_{L_{shift}}$	shift of sigmoidal microscopic benefit-of-leisure
$\delta(\cdot)$	delta/indicator function
$\mathbb{E}_\pi$	expected value with respect to policy π
K_L	slope of linear microscopic benefit-of-leisure
L	leisure
$\mu_a(\tau_a)$	effective prior probability density of choosing duration τ_a
P	Price
$\pi([a,\tau_a] \mid \vec{s})$	policy or choice rule: probability of choosing action a, for duration τ_a from state $\vec{s}$
post	post-reward
pre	pre-reward
$Q(\vec{s},[a,\tau_a])$	expected return or (differential) Q-value of taking action a, for duration τ_a from state $\vec{s}$
ρ	reward rate
$\rho\ \tau_a$	opportunity cost of time for taking action a for duration τ_a
RI	(subjective) Reward Intensity
$\frac{RI}{P}$	payoff
$\vec{s}$	state
TA	Time Allocation
τ_L	duration of instrumental leisure
τ_{Pav}	Pavlovian component of post-reward leisure
τ_W	duration of work
W	work
$w \in [0,P)$	amount of work time so far executed out of the price
$V(\vec{s})$	expected return or value of state $\vec{s}$

Symbol	Meaning
$1/\lambda$	mean of exponential effective prior probability density for leisure time. More explanation is here so we can see where more text would wrap.
$\alpha \in [0,1]$	weight on linear component of microscopic benefit-of-leisure
$\beta \in [0,\infty)$	inverse temperature or degree of stochasticity-determinism parameter
CHT	Cumulative Handling Time
$C_L(\cdot)$	microscopic benefit-of-leisure
$C_{L_{max}}$	maximum of sigmoidal microscopic benefit-of-leisure
$C_{L_{shift}}$	shift of sigmoidal microscopic benefit-of-leisure

Simple code:

```
procedure bubbleSort( A : list of sortable items )
    n = length(A)
    repeat
       newn = 0
       for i = 1 to n-1 inclusive do
          if A[i-1] > A[i] then
             swap(A[i-1], A[i])
             newn = i
          end if
       end for
       n = newn
    until n = 0
end procedure
```

Algorithm environment:

Algorithm 1 A sample in an algorithm environment.

```
if i ≥ maxval then
    i ← 0
else
    if i + k ≤ maxval then
        i ← i + k
    end if
end if
```

This decision has been studied by economists Battalio, Green, and Kagel (1981); Camerer, Babcock, Loewenstein, and Thaler (1997); Frank (2005); Green, Kagel, and Battalio (1987); Kagel, Battalio, and Green (1995), behavioural psychologists Baum (1974, 1981); Baum and Rachlin (1969); Dallery, McDowell, and Lancaster (2000); Green and Rachlin (1991); Herrnstein (1961, 1974); McDowell (1986, 2005); Skinner (1938, 1981), ethologists Haccou and Meelis (1992) and neuroscientists Arvanitogiannis and Shizgal (2008); Y.-A. Breton, Marcus, and Shizgal (2009); Conover and Shizgal (2005); Hernandez, Breton, Conover, and Shizgal (2010); Hernandez, Trujillo-Pisanty, Cossette, Conover, and Shizgal (2012); Niv, Daw, Joel, and Dayan (2007); Trujillo-Pisanty et al. (2011). Tasks involving free operant behaviour are particularly revealing, since subjects can choose what, when and how, minimally encumbered by direct experimenter intervention. We consider the cumulative handling time (CHT) schedule brain stimulation reward paradigm of Shizgal and colleagues Y.-A. Breton et al. (2009); Hernandez et al. (2010), in which animals have to invest quantifiable work to get rewards which are psychophysically stationary and repeatable.

Most previous investigations of time allocation have focused on *molar* or *macroscopic* characterisations of behaviour Baum (1976, 1995, 2001, 2002, 2004); Baum and Rachlin (1969); Camerer et al. (1997); Conover and Shizgal (2005); Frank (2005); Hernandez et al. (2010, 2012); Hineline (2001); Kagel et al. (1995); Rachlin (1978); Trujillo-Pisanty et al. (2011), capturing the average times allocated to work or leisure. Here, we characterize the detailed temporal topography of choice, i.e., the fine-scale *molecular* or *microscopic* structure of allocation Ferster and Skinner (1957); Gilbert (1958); Shull, Gaynor, and Grimes (2001); Williams, Sagvolden, Taylor, and Sagvolden (2009a, 2009b, 2009c), that is lost in molar averages (see Figure C). We build an approximately normative, reinforcement-learning, account, in which microscopic choices approximately maximize net benefit. Our central intent is to understand the qualitative structure of the molecular behaviour of subjects, providing an account that can generalize to many experimental paradigms. Therefore, although we apply the model to a set of CHT experiments in rats it is the next stage of the programme to fit this behaviour quantitatively in detail.

Having introduced previous approaches, we describe an example task and experiments (Section 2), key molecular features of the data from those (Section 3), our novel normative, microscopic approach (Section 4), and how it captures these key features (Section 5).

There are influential accounts from labour supply theory that have concentrated on the particularities of the macroscopic utility function that are necessary for the optimal allocation of time to include non-zero amounts of both work and leisure Camerer et al. (1997); Conover and Shizgal (2005); Frank (2005); Kagel et al. (1995). There have also been algorithmic accounts, parameterizing the allocation in order to elucidate the effects of manipulations such as drugs of addiction Hernandez et al. (2010, 2012); Trujillo-Pisanty et al. (2011).

RESULTS

Task and Experiment

As an example paradigm employed in rodents, consider a CHT task Y.-A. Breton et al. (2009); Hernandez et al. (2010) in which subjects choose between working–the facile task of holding down a light lever, and engaging in leisure, i.e., resting, grooming, exploring etc (Figure A). A brain stimulation reward (BSR; Olds and Milner (1954)) is given after the subject has accumulated work for an experimenter-defined total time-period called the *price* (P; see Table 1 for a description of all symbols). BSR does not suffer satiation and allows precise, psychophysically stable data to be collected over many months. We show data initially reported in Y. Breton, Conover, and Shizgal (2009) (and subsequently in Breton et al (in preparation), Solomon et al (in preparation)).

Furthermore, BSR has been shown to compete with Conover and Shizgal (1994a), summate with Conover and Shizgal (1994a, 1994b); Conover, Woodside, and Shizgal (1994), and substitute for Green and Rachlin (1991), gustatory rewards, demonstrating that these two at least partly share a common currency.

MATERIALS AND METHODS

The objective strength of the BSR is the frequency of electrical stimulation pulses applied to the medial forebrain bundle. This is assumed to have a subjective worth, or *microscopic utility* (to distinguish it from the *macroscopic utility* described in Arvanitogiannis and Shizgal (2008); Y.-A. Breton et al. (2009); Conover and Shizgal (2005); Hernandez et al. (2010, 2012); Trujillo-Pisanty et al. (2011)) called the *reward intensity* (RI, in arbitrary units). The transformation from objective to subjective worth has been previously determined Gallistel and Leon (1991); Hamilton, Stellar, and Hart (1985); Leon and Gallistel (1992); Mark and Gallistel (1993); Simmons and Gallistel (1994); Sonnenschein, Conover, and Shizgal (2003), The ratio of the reward intensity to the price is called the *payoff*. Leisure is assumed to have an intrinsic subjective worth, although its utility remains to be quantified. Throughout a task trial, the objective strength of the reward and price are held fixed. The total time the subjects could work per trial is 25 times the price (plus extra time for 'consuming' rewards). enabling at most 25 rewards to be harvested. A behaviourally observed work or leisure bout is defined as a temporally continuous act of working or engaging in leisure, respectively. Of course, contiguous short working or leisure bouts are externally indistinguishable from one long bout. Subjects are free to distribute leisure bouts in between individual work bouts.

Subjects face triads of trials: 'leading', 'test', then 'trailing' (Figure S1). Leading and trailing trials involve maximal and minimal reward intensities respectively, and the shortest price (we use the qualifiers "short", "long", etc. to emphasise that the price is an experimenter determined *time-period*). We analyze the sandwiched test trials, which span a range of prices and reward intensities. Leading and trailing trials allow calibration, so subjects can stably assess RI and P on test trials. Subjects tend to be at leisure on trailing trials, limiting physical fatigue. Subjects repeatedly experience each test reward intensity and price over many months, and so can readily appreciate them after minimal experience on a given trial without uncertainty.

In the future we hope molecular statistics will be derived from the detailed topography of depressing and releasing.

MOLAR AND MOLECULAR ANALYSES OF DATA

The key molar statistic is the Time Allocation (TA), namely the proportion of the available time for working in a test trial that the subject spends pressing the lever. Figure B shows example TAs for a typical subject. *TA* increases with the reward intensity and decreases with the price. Conversely, a molecular analysis, shown in the *ethograms* in (Figure C, D), assesses the detailed temporal topography of choice, recording when, and for how long, each act of work or leisure occurred (after the first acquisition of the reward in the trial, i.e., after the 'pink' lever presses in Figure D). The *TA* can be derived from the molecular ethogram data, but not vice-versa, since many different molecular patterns (Figure C) share a single *TA*.

Qualitative characteristics of the molecular structure of the data (Figure D) include:

Roman list:

(i) at high payoffs, subjects work almost continuously, engaging in little leisure in between work bouts;
(ii) at low payoffs, they engage in leisure all at once, in long bouts after working, rather than distributing the same amount of leisure time into multiple short leisure bouts;
(iii) subjects work continuously for the entire price duration, as long as the price is not very long (as shown by an analysis conducted by Y-AB, to be published separately);
(iv) the duration of leisure bouts is variable.

Numbered list:

1. at high payoffs, subjects work almost continuously, engaging in little leisure inbetween work bouts;
2. at low payoffs, they engage in leisure all at once, in long bouts after working, rather than distributing the same amount of leisure time into multiple short leisure bouts;
3. subjects work continuously for the entire price duration, as long as the price is not very long (as shown by an analysis conducted by Y-AB, to be published separately);
4. the duration of leisure bouts is variable.

Bulleted list:

- at high payoffs, subjects work almost continuously, engaging in little leisure inbetween work bouts;
- at low payoffs, they engage in leisure all at once, in long bouts after working, rather than distributing the same amount of leisure time into multiple short leisure bouts;
- subjects work continuously for the entire price duration, as long as the price is not very long (as shown by an analysis conducted by Y-AB, to be published separately);
- the duration of leisure bouts is variable.

SAMPLE EQUATIONS

$$\rho^\pi = \frac{RI + \mathbb{E}_{\pi([L,\tau_L]|\text{post})}\left[C_L(\tau_{\text{Pav}} + \tau_L)\right] + \int_0^P dw\, \mathbb{E}_{\pi_{w_L}}\left[\sum_{n_{L|[\text{pre},w]}} C_L(\tau_L)\right]}{P + \mathbb{E}_{\pi([L,\tau_L]|\text{post})}[\tau_L] + \tau_{\text{Pav}} + \int_0^P dw\, \mathbb{E}_{\pi_{w_L}}\left[\sum_{n_{L|[\text{pre},w]}} \tau_L\right]} \tag{1}$$

As long as $RI - K_L P > \frac{1}{\beta}$

$$\left.\begin{aligned} \rho^\pi &= \frac{\beta(RI + K_L\tau_{\text{Pav}}) - 1}{\beta(P + \tau_{\text{Pav}})} \\ \text{and} \qquad \mathbb{E}[\tau_L|\text{post}] &= \frac{P + \tau_{\text{Pav}}}{\beta(RI - K_L P) - 1} \end{aligned}\right\} \tag{2}$$

MICRO SEMI MARKOV DECISION PROCESS MODEL

We consider whether key features of the data in Figure D might arise from the subject's making stochastic optimal control choices, i.e., ones that at least approximately maximise the expected return arising from all benefits and costs over entire trials. Following Niv et al. (2007), we formulate this computational problem using the reinforcement learning framework of infinite horizon (Semi) Markov Decision Processes ((S)MDPs) Puterman (2005); Sutton and Barto (1998) (Figure 2A). Subjects not only choose which action *a* to take, i.e. to work (*W*) or engage in leisure (L), but also *the duration of the action* (τ_a). They pay an automatic *opportunity cost of time*: performing an action over a longer period denies the subject the opportunity to take other actions during that period, and thus extirpates any potential benefit from those actions.

Since trials are substantially extended, we assume the subjects do not worry about the time the trial ends, and instead make choices that would (approximately) maximize their average summed microscopic utility per unit time Niv et al. (2007). Nevertheless, for comparison with the data, we still terminate each trial at 25× price, so actions can be *censored* by the end of the trial, preventing their completion.

Utility The utility of the reward is RI. We assume that pressing the lever requires such minimal force that it does not incur any direct effort cost. We assume leisure to be intrinsically beneficial according to a function $C_L(\tau)$ of its duration (but formally independent of any other rewards or costs). The simplest such function is linear $C_L(\tau) = K_L\tau$ (Figure 2B, upper panel blue line), which would imply that the net utility of several short leisure bouts would be the same as a single bout of equal total length (Figure 2B, lower panel, blue line).

Alternatively, $C_L(\cdot)$ could be supra-linear (Figure 2B, upper panel, red curve). For this function, a single long leisure bout would be preferred to an equivalent time spent in several short bouts (Figure 2B, lower panel, red curve). If $C_L(\cdot)$ saturates, the rate of accrual of benefit-of-leisure $dC_L(\tau)/d\tau$ will peak at an optimal bout duration. We represent this class of functions with a sigmoid, although many other non-linearities are possible. Finally, to encompass both extremes, we consider a weighted sum of linear and sigmoid $C_L(\cdot)$, with the same maximal slope (Figure 2B, green curve. Linear $C_L(\cdot)$ has weight $\alpha = 1$, Eq. (S-3))

Evidence from related tasks Guitart-Masip et al. (2011); Shidara, Aigner, and Richmond (1998) suggests that the leisure time will be subject to Pavlovian as well as instrumental influences Breland and Breland (1961); Dayan, Niv, Seymour, and Daw (2006); Takikawa, Kawagoe, Itoh, Nakahara, and Hikosaka (2002). Subjects exhibit high error rates and slow reaction times for trials with high net payoffs, even when this is only detrimental. We formalize this with a leisure time as a sum of a mandatory Pavlovian contribution τ_{Pav} (in addition to the extra time for 'consuming' rewards), and an instrumental contribution τ_L, chosen, in the light of τ_{Pav}, to optimize the expected return. The Pavlovian component comprises a mandatory pause, which is curtailed by the subject's reengagement (conditioned-response) with the reward (unconditioned-stimulus)-predicting lever (conditioned-stimulus). As we shall discuss, we postulate a Pavlovian component to account for the detrimental leisure bouts at high payoffs. We assume $\tau_{\text{Pav}} = f_{\text{Pav}}(RI, P)$ decreases with payoff – i.e., increases with price and decreases with reward intensity (Figure 2C).

The net microscopic benefit-of-leisure is then $C_L(\tau_L + \tau_{\text{Pav}})$ over a bout of total length $\tau_L + \tau_{\text{Pav}}$.

We therefore refer to it as a micro SMDP. In the *Discussion* section we consider an alternate, nanoscopic variant which makes choices at a finer timescale.[1]

SAMPLE CITATIONS

`\citet{Hernandez2012}`: Hernandez et al. (2012) Textual citation;

`\citep{Hernandez2012}`: (Hernandez et al., 2012) Parenthetical citation;

`\citet*{Hernandez2012}`: Hernandez, Trujillo-Pisanty, Cossette, Conover, and Shizgal (2012) Same as `\citet` but if there are several authors, all names are printed;

`\citep*{Hernandez2012}`: (Hernandez, Trujillo-Pisanty, Cossette, Conover, & Shizgal, 2012) The same as `\citep` but if there are several authors, all names are printed;

`\citeauthor{Hernandez2012}`: Hernandez et al. Prints only the name of the authors(s);

[1] I would get rid of the following, since we aren't fitting behaviour quantitatively: Following previous experimental work Arvanitogiannis and Shizgal (2008); Y.-A. Breton et al. (2009); Hernandez et al. (2010, 2012); Trujillo-Pisanty et al. (2011), we have treated pre-reward leisure bouts which last less than 1 second as part of the previous work bout. Whether such "tapping" behaviour while working contradicts pre-commitment or indeed reflects lapses is an empirical question.

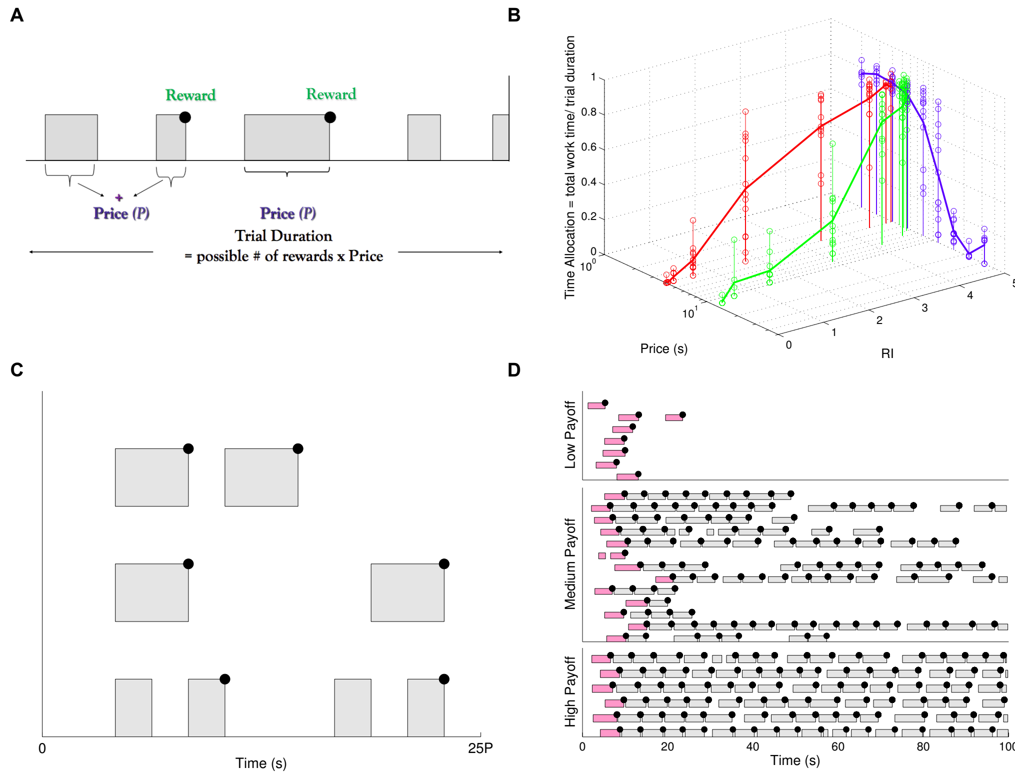

Figure 1. (Colour online) **Task and key features of the data.**
A) Cumulative handling time (CHT) task. Grey bars denote work (depressing a lever), white gaps show leisure. The subject must accumulate work up to a total period of time called the *price* (P) in order to obtain a single reward (black dot) of subjective reward intensity RI. The trial duration is $25 \times$ price (plus 2s each time the price is attained, during which the lever is retracted so it cannot work; not shown). The reward intensity and price are held fixed within a trial. B) Molar time allocation (TA) functions of a typical subject as a function of reward intensity and price. Red curves: effect of reward intensity, for a fixed short price; blue curves: effect of price, for a fixed high reward intensity; green curves: joint effect of reward intensity and price. C) A molecular analysis may reveal different microstructures of working and engaging in leisure. The three rows show three different hypothetical trials. All three microstructures have the same molar TA, but are clearly distinguishable. D) Molecular *ethogram* showing the detailed temporal topography of working and engaging in leisure for the subject in B). Upper, middle and lower panels show low, medium and high payoffs, respectively, for a fixed, short price. Following previous reports using rat subjects, releases shorter than 1 second are considered part of the previous work bout (since subjects remain at the lever during this period). *Graphically*, this makes some work bouts *appear* longer than others. The subject mostly pre-commits to working continuously for the entire price duration. When the payoff is high, the subject works almost continuously for the entire trial, engaging in very short leisure bouts inbetween work bouts. When the payoff is low, the subject engages in a long leisure bout after receiving a reward. This leisure bout is potentially longer than the trial, whence it would be censored. The part of a trial before the reward, price and probability of reward delivery are certainly known is coloured pink and not considered further. Data collected by Y-AB and RS and initially reported in Y. Breton et al. (2009).

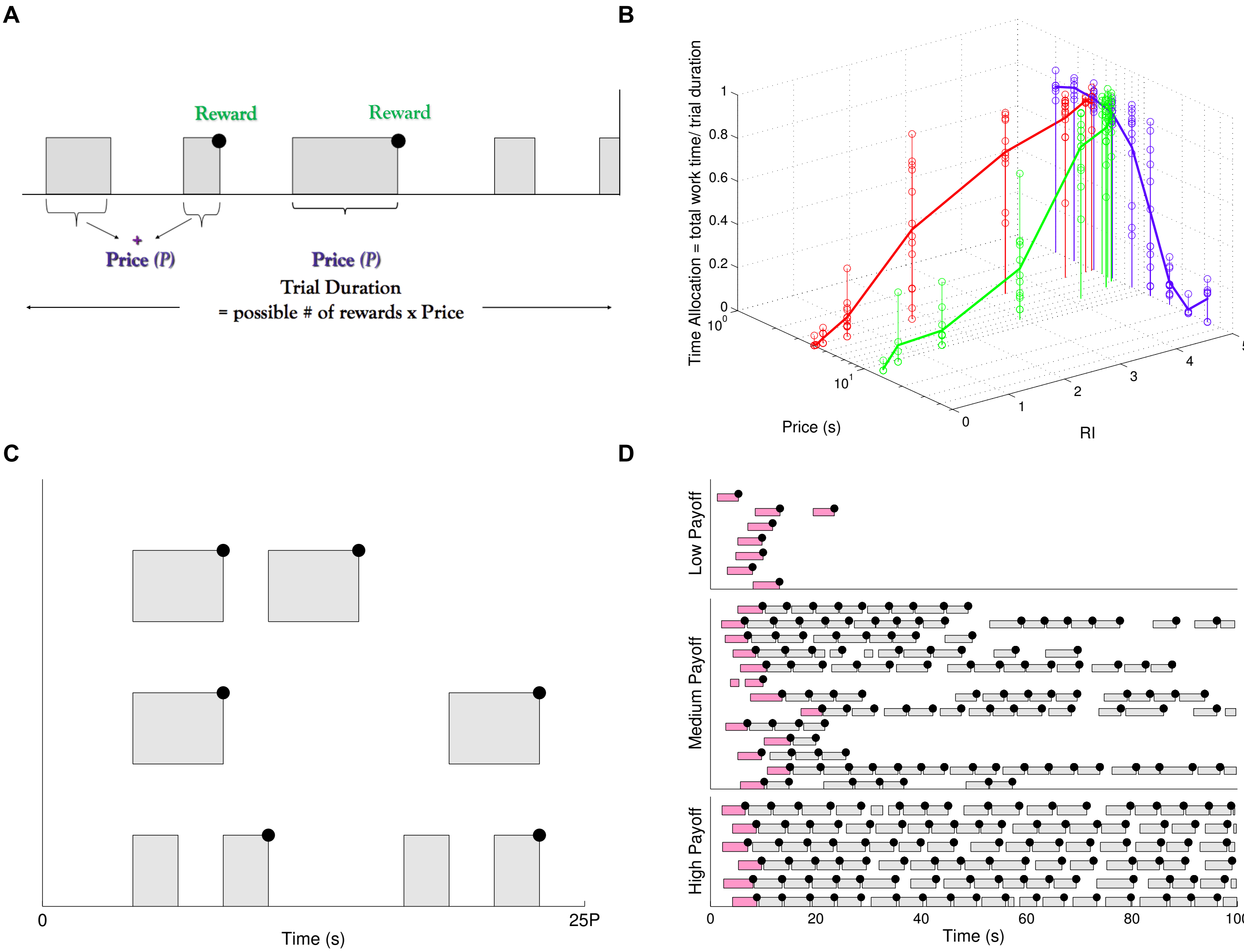

Figure 2. (Colour online) **Task and key features of the data.**
A) Cumulative handling time (CHT) task. Grey bars denote work (depressing a lever), white gaps show leisure. The subject must accumulate work up to a total period of time called the *price* (P) in order to obtain a single reward (black dot) of subjective reward intensity RI. The trial duration is 25 × price (plus 2s each time the price is attained, during which the lever is retracted so it cannot work; not shown). The reward intensity and price are held fixed within a trial. B) Molar time allocation (TA) functions of a typical subject as a function of reward intensity and price. Red curves: effect of reward intensity, for a fixed short price; blue curves: effect of price, for a fixed high reward intensity; green curves: joint effect of reward intensity and price. C) A molecular analysis may reveal different microstructures of working and engaging in leisure. The three rows show three different hypothetical trials. All three microstructures have the same molar TA, but are clearly distinguishable. D) Molecular *ethogram* showing the detailed temporal topography of working and engaging in leisure for the subject in B). Upper, middle and lower panels show low, medium and high payoffs, respectively, for a fixed, short price. Following previous reports using rat subjects, releases shorter than 1 second are considered part of the previous work bout (since subjects remain at the lever during this period). *Graphically*, this makes some work bouts *appear* longer than others. The subject mostly pre-commits to working continuously for the entire price duration. When the payoff is high, the subject works almost continuously for the entire trial, engaging in very short leisure bouts inbetween work bouts. When the payoff is low, the subject engages in a long leisure bout after receiving a reward. Data collected by Y-AB and RS and initially reported in Y. Breton et al. (2009).

Table 1. Time of the Transition Between Phase 1 and Phase 2[a]

Run	Time (min)
$l1$	260
$l2$	300
$l3$	340
$h1$	270
$h2$	250
$h3$	380
$r1$	370
$r2$	390

[a]Table note text here.

Table 2. Sample table taken from [treu03]

POS	chip	ID	X	Y	RA	DEC	IAU$\pm$ δ IAU	IAP1$\pm$ δ IAP1	IAP2 $\pm$ δ IAP2	star	E	Comment
0	2	1	1370.99	57.35[a]	6.651120	17.131149	21.344$\pm$0.006[b]	2 4.385$\pm$0.016	23.528$\pm$0.013	0.0	9	-
0	2	2	1476.62	8.03	6.651480	17.129572	21.641$\pm$0.005	2 3.141$\pm$0.007	22.007$\pm$0.004	0.0	9	-
0	2	3	1079.62	28.92	6.652430	17.135000	23.953$\pm$0.030	2 4.890$\pm$0.023	24.240$\pm$0.023	0.0	-	-
0	2	4	114.58	21.22	6.655560	17.148020	23.801$\pm$0.025	2 5.039$\pm$0.026	24.112$\pm$0.021	0.0	-	-
0	2	5	46.78	19.46	6.655800	17.148932	23.012$\pm$0.012	2 3.924$\pm$0.012	23.282$\pm$0.011	0.0	-	-
0	2	6	1441.84	16.16	6.651480	17.130072	24.393$\pm$0.045	2 6.099$\pm$0.062	25.119$\pm$0.049	0.0	-	-
0	2	7	205.43	3.96	6.655520	17.146742	24.424$\pm$0.032	2 5.028$\pm$0.025	24.597$\pm$0.027	0.0	-	-
0	2	8	1321.63	9.76	6.651950	17.131672	22.189$\pm$0.011	2 4.743$\pm$0.021	23.298$\pm$0.011	0.0	4	edge

Table 2 is published in its entirety in the electronic edition of the *Astrophysical Journal*.

[a] Sample footnote for table 2.

[b] Another sample footnote for table 2.

Table 3. Here is a caption for a table that is found in landscape mode.

POS	chip	ID	X	Y	RA	DEC	IAU± δ IAU	IAP1± δ IAP1	IAP2 ± δ IAP2	star	E	Comment
0	2	1	1370.99	57.35[a]	6.651120	17.131149	21.344±0.006[b]	2 4.385±0.016	23.528±0.013	0.0	9	-
0	2	2	1476.62	8.03	6.651480	17.129572	21.641±0.005	2 3.141±0.007	22.007±0.004	0.0	9	-
0	2	3	1079.62	28.92	6.652430	17.135000	23.953±0.030	2 4.890±0.023	24.240±0.023	0.0	-	-
0	2	4	114.58	21.22	6.655560	17.148020	23.801±0.025	2 5.039±0.026	24.112±0.021	0.0	-	-
0	2	5	46.78	19.46	6.655800	17.148932	23.012±0.012	2 3.924±0.012	23.282±0.011	0.0	-	-
0	2	6	1441.84	16.16	6.651480	17.130072	24.393±0.045	2 6.099±0.062	25.119±0.049	0.0	-	-
0	2	7	205.43	3.96	6.655520	17.146742	24.424±0.032	2 5.028±0.025	24.597±0.027	0.0	-	-
0	2	8	1321.63	9.76	6.651950	17.131672	22.189±0.011	2 4.743±0.021	23.298±0.011	0.0	4	edge

Table 2 is published in its entirety in the electronic edition of the *Astrophysical Journal*.

[a] Sample footnote for table 2.

[b] Another sample footnote for table 2.

DISCUSSION

Real time decision-making involves choices about when and for how long to execute actions as well as well as which to perform. We studied a simplified version of this problem, considering a paradigmatic case with economic, psychological, ethological and biological consequences, namely working for explicit external rewards versus engaging in leisure for its own implicit benefit. We offered a normative, microscopic framework accounting for subjects' temporal choices, showing the rich collection of effects associated with the way that the subjective benefit-of-leisure grows with its duration.

Our microscopic formulation involved an infinite horizon Semi-Markov Decision Process (SMDP) with three key characteristics: approximate optimization of the reward rate, stochastic choices as a function of the values of the options concerned, and an assumption that, a priori, temporal choices would never be infinitely extended (owing to either lapses or the greater uncertainty that accompanies the timing of longer intervals Gibbon (1977)). The metrics associated with this last assumption had little effect on the output of the model. We may have alternately assumed that arbitrarily long durations could be chosen as frequently as short ones, but more noisily executed; we imputed all such noise to the choice rule for simplicity.

We exercised our model by examining a psychophysical paradigm called the cumulative handling time (CHT) schedule involving brain stimulation reward.

The CHT controls both the (average) minimum inter-reward interval and the amount of work required to earn a reward. More common schedules of reinforcement such as Fixed Ratio, or Variable Interval control one but not the other. This makes the CHT particularly useful for studying the choice of how long to either work or engage in leisure.

Table 4: ApJ costs from 1991 to 2013

Year	Subscription cost ($)	Publication charges ($/page)
1991	600	100
1992	650	105
1993	550	103
1994	450	110
1995	410	112
1996	400	114
1997	525	115
1998	590	116
1999	575	115
2000	450	103
2001	490	90
2002	500	88
2003	450	90
2004	460	88

Table continued on next page

Table 4, *continued from previous page.*

ApJ costs from 1991 to 2013

Year	Subscription cost ($)	Publication charges ($/page)
2005	440	79
2006	350	77
2007	325	70
2008	320	65
2009	190	68
2010	280	70
2011	275	68
2012	150	56
2013	140	55

Nevertheless, it would be straightforward to adapt our model to treat waiting schedules such as J. Bizot, Le Bihan, Puech, Hamon, and Thiébot (1999); J. C. Bizot, Thiébot, Le Bihan, Soubri�, and Simon (1988); Fletcher (1995); Ho, Al-Zahrani, Al-Ruwaitea, Bradshaw, and Szabadi (1998); Jolly, Richards, and Seiden (1999); K. Miyazaki, Miyazaki, and Doya (2011); K. W. Miyazaki, Miyazaki, and Doya (2012) or to add other facets. For instance, effort costs would lead to shorter work bouts rather than the pre-commitment to working for the duration of the price observed in the data. Costs of waiting through a delay would also lead subjects to quit waiting earlier than later. Other tasks with other work requirements could also be fitted into the model by changing the state and transition structure of the Markov chain. The main issue the CHT task poses for the model is that it is separated into episodic trials of different types making infinite horizon optimization an approximation. However, the approximation is likely benign, since the relevant trials are extended (each lasts 25 times the price), and the main effect is that work and leisure bouts can sometimes be censored at the ends of trials.

SUPPORTIVE INFORMATION

No supportive information is available at this time.

ACKNOWLEDGMENTS

The authors thank Laurence Aitchison for fruitful discussions. RKN and PD received funding from the Gatsby Charitable Foundation. Y-AB, RBS, KC and PS received funding from Canadian Institutes of Health Research grant $MOP74577$, Fond de recherche Québec - Santé (Group grant to the Groupe de recherche en neurobiologie comportementale, Shimon Amir, P.I.), and Concordia University Research Chair (Tier I).

AUTHOR CONTRIBUTIONS

Project was formulated by RKN, PD, PS, based on substantial data, analyses and experiments of Y-AB, KC, RS, PS. RKN, PD formalised the model, RKN implemented and ran the model; RKN analysed the molecular ethogram data; Y-AB formalised and implemented a CTMC model. All authors wrote the manuscript.

REFERENCES

Arvanitogiannis, A., & Shizgal, P. (2008). The reinforcement mountain: allocation of behavior as a function of the rate and intensity of rewarding brain stimulation. *Behav Neurosci*, *122*(5), 1126–38. doi: 10.1037/a0012679

Battalio, R. C., Green, L., & Kagel, J. H. (1981). Income-Leisure Tradeoffs of Animal Workers. *The American Economic Review*, *71*(4), 621–632.

Baum, W. M. (1974). On two types of deviation from the matching law: bias and undermatching. *J Exp Anal Behav*, *22*(1), 231–42.

Baum, W. M. (1976). Time-based and count-based measurement of preference. *J Exp Anal Behav*, *26*(1), 27–35.

Baum, W. M. (1981). Optimization and the matching law as accounts of instrumental behavior. *J Exp Anal Behav*, *36*(3), 387–403.

Baum, W. M. (1995). Introduction to molar behavior analysis. *Mexican Journal of Behavior Analysis*, *21*, 7–25.

Baum, W. M. (2001). Molar versus as a paradigm clash. *J Exp Anal Behav*, *75*(3), 338–41; discussion 367–78.

Baum, W. M. (2002). From molecular to molar: a paradigm shift in behavior analysis. *J Exp Anal Behav*, *78*(1), 95–116.

Baum, W. M. (2004). Molar and molecular views of choice. *Behavioural Processes*, *66*(3), 349–59.

Baum, W. M., & Rachlin, H. C. (1969). Choice as time allocation. *J Exp Anal Behav*, *12*(6), 861–74.

Bizot, J., Le Bihan, C., Puech, A. J., Hamon, M., & Thiébot, M. (1999). Serotonin and tolerance to delay of reward in rats. *Psychopharmacology (Berl)*, *146*(4), 400–412.

Bizot, J. C., Thiébot, M. H., Le Bihan, C., Soubri�, P., & Simon, P. (1988). Effects of imipramine-like drugs and serotonin uptake blockers on delay of reward in rats. possible implication in the behavioral mechanism of action of antidepressants. *J Pharmacol Exp Ther*, *246*(3), 1144–1151.

Breland, K., & Breland, M. (1961). The misbehavior of organisms. *Am Psychol*, *16*(11), 681–684. doi: 10.1037/h0040090

Breton, Y., Conover, K., & Shizgal, P. (2009). Probability discounting of brain stimulation reward in the rat.. (39th Annual Meeting of the Society for Neuroscience (Neuroscience 2009))

Breton, Y.-A., Marcus, J. C., & Shizgal, P. (2009). Rattus Psychologicus: construction of preferences by self-stimulating rats. *Behav Brain Res*, *202*(1), 77–91. doi: 10.1016/j.bbr.2009.03.019

Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997). Labor Supply of New York City Cabdrivers: One Day at a Time. *Q J Econ*, *112*(2), 407–441. doi: 10.1162/003355397555244

Conover, K. L., & Shizgal, P. (1994a). Competition and summation between rewarding effects of sucrose and lateral hypothalamic stimulation in the rat. *Behav Neurosci*, *108*(3), 537–48.

Conover, K. L., & Shizgal, P. (1994b). Differential effects of postingestive feedback on the reward value of sucrose and lateral hypothalamic stimulation in rats. *Behav Neurosci*, *108*(3), 559–72.

Conover, K. L., & Shizgal, P. (2005). Employing labor-supply theory to measure the reward value of electrical brain stimulation. *Games and Economic Behavior*, *52*(2), 283–304. doi: 10.1016/j.geb.2004.08.003

Conover, K. L., Woodside, B., & Shizgal, P. (1994). Effects of sodium depletion on competition and summation between rewarding effects of salt and lateral hypothalamic stimulation in the rat. *Behav Neurosci*, *108*(3), 549–58.

Dallery, J., McDowell, J. J., & Lancaster, J. S. (2000). Falsification of matching theory's account of single-alternative responding: Herrnstein's k varies with sucrose concentration. *J Exp Anal Behav*, *73*(1), 23–43. doi: 10.1901/jeab.2000.73-23

Dayan, P., Niv, Y., Seymour, B., & Daw, N. D. (2006). The misbehavior of value and the discipline of the will. *Neural Net*, *19*(8), 1153–60. doi: 10.1016/j.neunet.2006.03.002

Ferster, C., & Skinner, B. F. (1957). *Schedules of reinforcement*. New York: Appleton-Century-Crofts.

Fletcher, P. J. (1995). Effects of combined or separate 5,7-dihydroxytryptamine lesions of the dorsal and median raphe nuclei on responding maintained by a drl 20s schedule of food reinforcement. *Brain Res*, *675*(1-2), 45–54.

Frank, R. H. (2005). *Microeconomics and Behavior*. New York: McGraw-Hill Higher Education.

Gallistel, C. R., & Leon, M. (1991). Measuring the subjective magnitude of brain stimulation reward by titration with rate of reward. *Behav Neurosci*, *105*(6), 913–25.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psych Rev*, *84*(3), 279–325.

Gilbert, T. F. (1958). Fundamental dimensional properties of the operant. *Psych Rev*, *65*(5), 272–82.

Green, L., Kagel, J. H., & Battalio, R. C. (1987). Consumption-leisure tradeoffs in pigeons: Effects of changing marginal wage rates by varying amount of reinforcement. *J Exp Anal Behav*, *47*(1), 17–28.

Green, L., & Rachlin, H. (1991). Economic substitutability of electrical brain stimulation, food, and water. *J Exp Anal Behav*, *55*(2), 133–43. doi: 10.1901/jeab.1991.55-133

Guitart-Masip, M., Fuentemilla, L., Bach, D. R., Huys, Q. J. M., Dayan, P., Dolan, R. J., & Duzel, E. (2011). Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. *J Neurosci*, *31*(21), 7867–75. doi: 10.1523/JNEUROSCI.6376-10.2011

Haccou, P., & Meelis, E. (1992). *Statistical Analysis of Behavioural Data: An Approach Based on Time-structured Models*. Oxford University Press, USA.

Hamilton, A. L., Stellar, J. R., & Hart, E. B. (1985). Reward, performance, and the response strength method in self-stimulating rats: validation and neuroleptics. *Phys & Behav*, *35*(6), 897–904.

Hernandez, G., Breton, Y.-A., Conover, K., & Shizgal, P. (2010). At what stage of neural processing does cocaine act to boost pursuit of rewards? *PloS one*, *5*(11), e15081. doi: 10.1371/journal.pone.0015081

Hernandez, G., Trujillo-Pisanty, I., Cossette, M.-P., Conover, K., & Shizgal, P. (2012). Role of Dopamine Tone in the Pursuit of Brain Stimulation Reward. *J Neurosci*, *32*(32), 11032–11041.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *J Exp Anal Behav*, *4*, 267–72. doi: 10.1901/jeab.1961.4-267

Herrnstein, R. J. (1974). Formal properties of the matching law. *J Exp Anal Behav*, *21*(1), 159–64.

Hineline, P. N. (2001). Beyond the molar-molecular distinction: we need multiscaled analyses. *J Exp Anal Behav*, *75*(3), 342–7; discussion 367–78. doi: 10.1901/jeab.2001.75-342

Ho, M. Y., Al-Zahrani, S. S., Al-Ruwaitea, A. S., Bradshaw, C. M., & Szabadi, E. (1998). 5-hydroxytryptamine and impulse control: prospects for a behavioural analysis. *J Psychopharmacol*, *12*(1), 68–78.

Jolly, D. C., Richards, J. B., & Seiden, L. S. (1999). Serotonergic mediation of drl 72s behavior: receptor subtype involvement in a behavioral screen for antidepressant drugs. *Biol Psychiatry*, *45*(9), 1151–1162.

Kagel, J. H., Battalio, R. C., & Green, L. (1995). *Economic Choice Theory: An Experimental Analysis of Animal Behavior*. Cambridge University Press.

Leon, M., & Gallistel, C. R. (1992). The function relating the subjective magnitude of brain stimulation reward to stimulation strength varies with site of stimulation. *Behav Brain Res*, *52*(2), 183–93.

Mark, T. A., & Gallistel, C. R. (1993). Subjective reward magnitude of medial forebrain stimulation as a function of train duration and pulse frequency. *Behav Neurosci*, *107*(2), 389–401.

McDowell, J. J. (1986). On the falsifiability of matching theory. *J Exp Anal Behav*, *45*(1), 63–74. doi: 10.1901/jeab.1986.45-63

McDowell, J. J. (2005). On the classic and modern theories of matching. *J Exp Anal Behav*, *84*(1), 111–27. doi: 10.1901/jeab.2005.59-04

Miyazaki, K., Miyazaki, K. W., & Doya, K. (2011). Activation of dorsal raphe serotonin neurons underlies waiting for delayed rewards. *J Neurosci*, *31*(2), 469–79. doi: 10.1523/JNEUROSCI.3714-10.2011

Miyazaki, K. W., Miyazaki, K., & Doya, K. (2012). Activation of dorsal raphe serotonin neurons is necessary for waiting for delayed rewards. *J Neurosci*, *32*(31), 10451–7. doi: 10.1523/JNEUROSCI.0915-12.2012

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507–20. doi: 10.1007/s00213-006-0502-4

Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J Comp and Phys Psych*, *47*(6), 419–27.

Puterman, M. L. (2005). *Markov Decision Processes: Discrete Stochastic Dynamic Programming (Wiley Series in Probability and Statistics)*. Wiley-Blackwell.

Rachlin, H. (1978). A molar theory of reinforcement schedules. *J Exp Anal Behav*, *30*(3), 345–60.

Shidara, M., Aigner, T. G., & Richmond, B. J. (1998). Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *J Neurosci*, *18*(7), 2613–25.

Shull, R. L., Gaynor, S. T., & Grimes, J. A. (2001). Response rate viewed as engagement bouts: effects of relative reinforcement and schedule type. *J Exp Anal Behav*, *75*(3), 247–74. doi: 10.1901/jeab.2001.75-247

Simmons, J. M., & Gallistel, C. R. (1994). Saturation of subjective reward magnitude as a function of current and pulse frequency. *Behav Neurosci*, *108*(1), 151–60.

Skinner, B. F. (1938). *The behavior of organisms: an experimental analysis*. New York: Appleton-Century-Crofts.

Skinner, B. F. (1981). Selection by consequences. *Science (New York, N.Y.)*, *213*(4507), 501–4.

Sonnenschein, B., Conover, K., & Shizgal, P. (2003). Growth of brain stimulation reward as a function of duration and stimulation strength. *Behav Neurosci*, *117*(5), 978–94. doi: 10.1037/0735-7044.117.5.978

Sutton, R., & Barto, A. (1998). *Reinforcement learning: An introduction* (Vol. 28). Cambridge University Press.

Takikawa, Y., Kawagoe, R., Itoh, H., Nakahara, H., & Hikosaka, O. (2002). Modulation of saccadic eye movements by predicted reward outcome. *Exp Brain Res*, *142*(2), 284–91. doi: 10.1007/s00221-001-0928-1

Trujillo-Pisanty, I., Hernandez, G., Moreau-Debord, I., Cossette, M.-P., Conover, K., Cheer, J. F., & Shizgal, P. (2011). Cannabinoid receptor blockade reduces the opportunity cost at which rats maintain operant performance for rewarding brain stimulation. *J Neurosci*, *31*(14), 5426–35. doi: 10.1523/JNEUROSCI.0079-11.2011

Williams, J., Sagvolden, G., Taylor, E., & Sagvolden, T. (2009a). Dynamic behavioural changes in the Spontaneously Hyperactive Rat: 1. Control by place, timing, and reinforcement rate. *Behav Brain Res*, *198*(2), 273–82. doi: 10.1016/j.bbr.2008.08.044

Williams, J., Sagvolden, G., Taylor, E., & Sagvolden, T. (2009b). Dynamic behavioural changes in the Spontaneously Hyperactive Rat: 2. Control by novelty. *Behav Brain Res*, *198*(2), 283–90. doi: 10.1016/j.bbr.2008.08.045

Williams, J., Sagvolden, G., Taylor, E., & Sagvolden, T. (2009c). Dynamic behavioural changes in the Spontaneously Hyperactive Rat: 3. Control by reinforcer rate changes and predictability. *Behav Brain Res*, *198*(2), 291–7. doi: 10.1016/j.bbr.2008.08.046

A: APPENDIX

We derive the result in Eq. (B.1). We consider a linear $C_L(\tau_L + \tau_{\text{Pav}}) = K_L(\tau_L + \tau_{\text{Pav}})$, and make two further simplifications: (i) the subject does not engage in leisure in the pre-reward state (and so works for the whole price when it works); and (ii) *a priori*, arbitrarily long leisure durations are possible ($\lambda = 0$). Then the reward rate in Eq. (1) becomes

$$\rho^\pi = \frac{RI + K_L\{\,\mathbb{E}[\tau_L|\text{post}] + \tau_{\text{Pav}}\}}{P + \mathbb{E}[\tau_L|\text{post}] + \tau_{\text{Pav}}} \tag{A.1}$$

As discussed in the *Results* section, the probability of engaging in instrumental leisure in the post-reward state is $\pi([L,\tau_L]\,|\text{post}) = \exp\left[-\{\beta(\rho^\pi - K_L)\}\tau_L\right]$, which is an exponential distribution with mean

$$\mathbb{E}[\tau_L|\text{post}] = \frac{1}{\beta(\rho^\pi - K_L)} \tag{A.2}$$

Re-arranging terms of this equation,

$$\rho^\pi = \frac{1}{\beta\,\mathbb{E}[\tau_L|\text{post}]} + K_L \tag{A.3}$$

Equating Eqs. (A.1) and (A.3) and solving for the mean instrumental leisure duration $\mathbb{E}[\tau_L|\text{post}]$, we derive

$$\mathbb{E}[\tau_L|\text{post}] = \frac{P + \tau_{\text{Pav}}}{\beta(RI - K_L P) - 1} \tag{A.4}$$

which is the second line of Eq.(B.1). This is the mean instrumental leisure duration as long as $RI - K_L P > 1$, and $\mathbb{E}[\tau_L|\text{post}] \to \infty$ otherwise. When the former condition holds, we may substitute Eq. (A.4) into Eq. (A.1) and solve for ρ^π

$$\begin{aligned}\rho^\pi &= \frac{(RI - K_L P)\ \ [\beta(RI + K_L\tau_{\text{Pav}}) - 1]}{(RI - K_L P)\ \beta(P + \tau_{\text{Pav}})}\\ &= \frac{\beta(RI + K_L\tau_{\text{Pav}}) - 1}{\beta(P + \tau_{\text{Pav}})}\end{aligned} \tag{A.5}$$

which is the first line of Eq.(B.1).

B: EXPECTED REWARD

Then the expected reward rate and mean leisure duration can be derived analytically (see Appendix). As long as $RI - K_L P > \frac{1}{\beta}$ As long as $pRI - K_L P - 2c_{dec} > \frac{1}{\beta}$

$$\begin{aligned} \rho^{\pi} &= \frac{\beta(RI + K_L \tau_{\text{Pav}}) - 1}{\beta(P + \tau_{\text{Pav}})} \\ \mathbb{E}[\tau_L|\text{post}] &= \frac{P + \tau_{\text{Pav}}}{\beta(RI - K_L P) - 1} \end{aligned} \tag{B.1}$$

Sample itemized list in appendix

- at high payoffs, subjects work almost continuously, engaging in little leisure inbetween work bouts;
- at low payoffs, they engage in leisure all at once, in long bouts after working, rather than distributing the same amount of leisure time into multiple short leisure bouts;
- subjects work continuously for the entire price duration, as long as the price is not very long (as shown by an analysis conducted by Y-AB, to be published separately);
- the duration of leisure bouts is variable.