## Appendix B.
## A formal description of the decision problem

This appendix offers a formal description of the decision problem that an agent faces. Appendix C describes how we find the optimal policy.

Figure B.1. provides a schematic overview of the decision problem. An agent's state has two components: a soma, representing the agent's condition, whether it be physical, social, or material; and belief, representing the agent's knowledge about the resources in its environment. An agent's somatic state can range from 0 to 100 (at 0, the agent dies). An agent changes this state as it interacts with its environment via resource encounters and extrinsic events. A resource encounter is always followed by an extrinsic event, and vice versa. We call the pairing of a resource encounter and the subsequent extrinsic event a cycle. As long as an agent is alive, there is a next cycle.

The resource quality, extrinsic event quality, the costs of sampling a cue (see below) and the somatic state, have discrete values that are multiples of 0.2. Hence the somatic state is a multiple of 0.2. Let $B$ denote the set of all possible somatic states:

$$B = \{ x \mid x = 0.2n, n \in \mathbb{N}_0, 0 \leq x \leq 100\} \quad \text{(B.1)}$$

Where $\mathbb{N}_0$ is the set of all natural numbers including 0. (In appendix C we use S to refer to the combination of an agent's somatic and belief state. To avoid confusion, we use B, short for '*budget*', to refer specifically to the somatic state). Let $b_t$ denote the somatic state at the start of cycle $t$.

### 1.1. Encountering a resource

The resource quality ranges between -20 and +20 and is a multiple of 0.2. Let $Q$ denote the set of all possible resource qualities:

$$Q = \{ x \mid x = 0.2n, n \in \mathbb{N}_0, -20 \leq x \leq 20\} \quad \text{(B.2)}$$

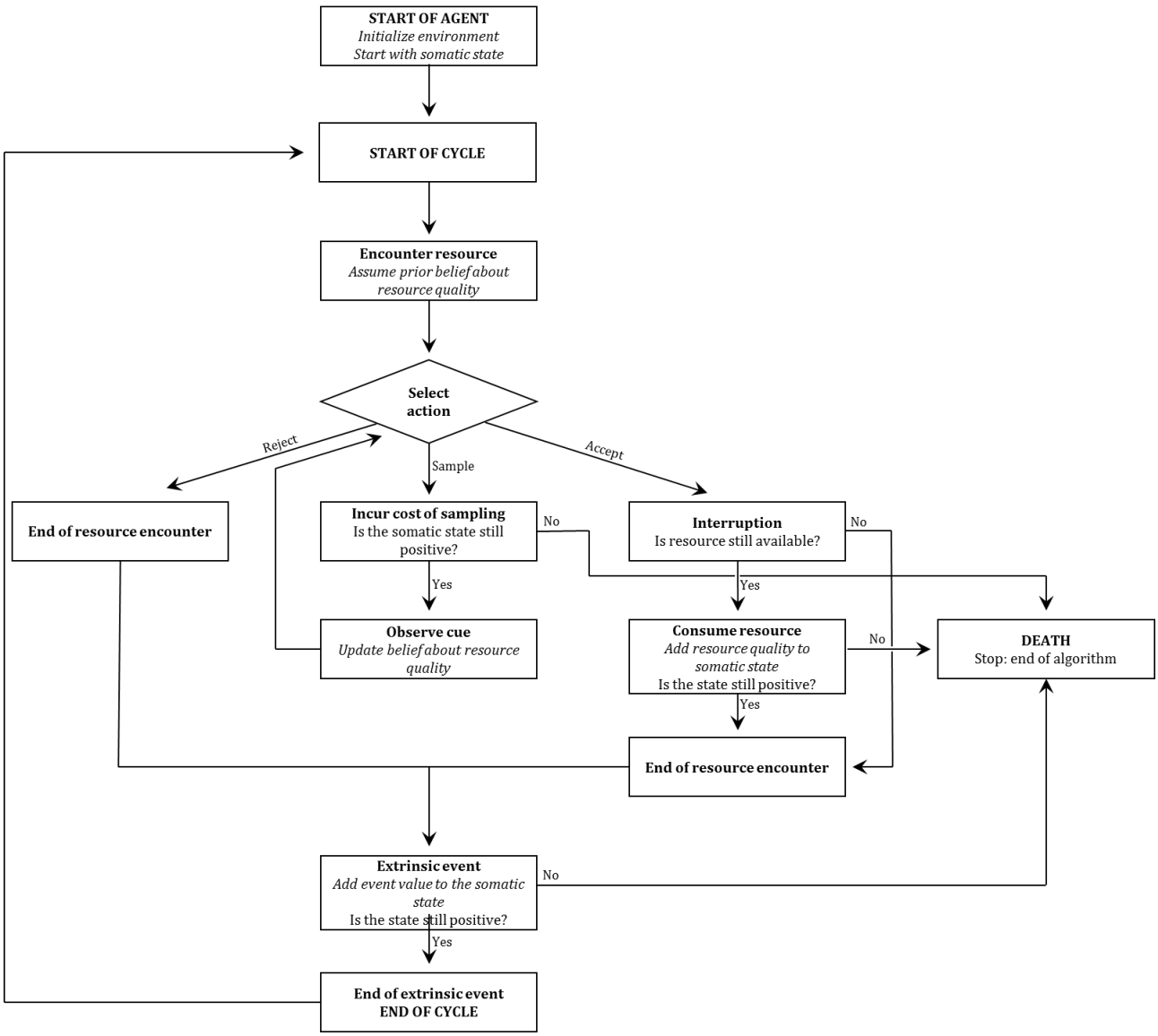


**Fig. B.1.** A schematic overview of the decision problem an agent faces.

Resources are independently and identically distributed and drawn from a truncated normal distribution with mean $\mu_{resource}$ and standard deviation $\sigma_{resource}$. Let $\eta(q|\,\mu_{resource}, \sigma_{resource})$ be the probability that a resource has quality $q$ (i.e., the probability mass function). The probability of a specific resource quality $q \in Q$ is:

$$Pr(q|\mu_{resource}, \sigma_{resource}) = \frac{\eta(q|\,\mu_{resource}, \sigma_{resource})}{\sum_{r \in Q} \eta(r|\,\mu_{resource}, \sigma_{resource})} \tag{B.3}$$

The denominator functions as a renormalizing constant, ensuring that all probabilities sum to 1.

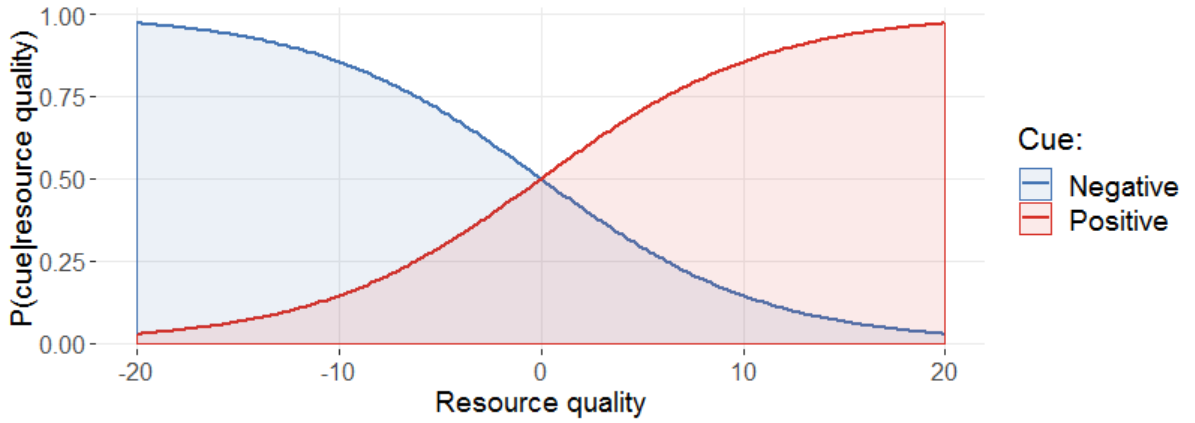


**Fig. B.2.** The normalized cue emission probabilities: the probability of observing a positive or negative cue given the resource quality. The cue reliability for positive cues is a normal distribution with a mean of 20; for negative cues this mean is -20. For both the standard deviation is determined by *ca*, which we have set to 15.

**1.1.1. Sampling cues.** An agent does not know the resource quality before accepting or rejecting. However, before making a decision, an agent can repeatedly sample a cue (up to a maximum of 10). Sampling, however, is costly: each sampled cue reduces an agent's somatic state by a fixed value, $c$. We set $c$ to 0.2.

A cue is either positive or negative ($c_+$ and $c_-$). A cue provides imperfect information about the resource quality; that is, cues reduce uncertainty, but do not eliminate it. The probability of observing a positive or negative cue depends on the *cue reliability,* $\Pr(c_+|q)$ and $\Pr(c_-|q)$, respectively. The cue reliability depends both on the resource quality and on the cue accuracy. When a resource has a positive quality, a positive cue is more likely than a negative cue. When it has a low quality, a negative cue is more likely (figure B.2). The cue accuracy, $ca$, determines how much more likely a positive (negative) resource results in positive (negative) cue. As $ca$ approaches 0, a positive (negative) resource deterministically produces positive (negative) cues. Conversely, as $ca$ approaches infinity, positive and negative cues become equally likely, independent of the resource quality. We set $ca$ to 15.

The non-normalized probability of observing a negative cue when a resource has quality $q$ is a truncated normal distribution with a mean of -20 and standard deviation $ca$:

$$\mathrm{N}(c\_|q) = \exp\left(-\frac{(q+20)^2}{2ca^2}\right) \tag{B.4}$$

Similarly, the non-normalized probability of observing a positive cue when the resource has quality $q$ is drawn from a distribution with mean 20 and standard deviation $ca$:

$$\mathrm{N}(c_+|q) = \exp\left(-\frac{(q-20)^2}{2ca^2}\right) \tag{B.5}$$

As a cue is either positive or negative, the probability of a positive and a negative cue must sum to 1. To ensure this, both probabilities are divided by a normalizing constant:

$$\Pr(c\_|q) = \frac{\mathrm{N}(\mathrm{c}\_|q)}{\mathrm{N}(c\_|q) + \mathrm{N}(c_+|q)} \tag{B.6}$$

$$\Pr(c_+|q) = \frac{\mathrm{N}(\mathrm{c}_+|q)}{\mathrm{N}(c\_|q) + \mathrm{N}(c_+|q)} \tag{B.7}$$

**1.1.2. Posterior beliefs**. After sampling a cue, an agent updates its belief about the resource quality using Bayes theorem. Let $B(q|D)$ denote this posterior belief. Let $D$ denote the set of $n_{c+}$ sampled positive cues and $n_{c-}$ the set of sampled negative cues.

$$\mathrm{B}(q|D) = \frac{\Pr(D|q) * \mathrm{B}(q|\emptyset)}{\sum_{r\in Q}\Pr(D|r) * \mathrm{B}(r|\emptyset)} \tag{B.8}$$

Where $\Pr(D|q)$ is the probability of obtaining $D$ given that the resource has quality $q$. We assume that evolution by natural selection has shaped agents to have prior beliefs that match the true distribution of resource quality in its (stationary) environment. Let $B(q|\emptyset)$ denote this *prior,* i.e., the belief of an agent before having sampled any cues:

$$B(q|\emptyset) = \Pr\left(q\middle|\mu_{resource}, \sigma_{resource}\right) \tag{B.9}$$

The probability of $D$ given $q$ is the conditional probability of observing $n_{c+}$ positive cues and $n_{c-}$ negative cues given $q$. As the binomial distribution specifies, the order in which cues are sampled is irrelevant:

$$\Pr(D|q) = \binom{n_{c+} + n_{c-}}{n_{c+}} \Pr(c_-|q)^{n_{c-}} \Pr(c_+|q)^{n_{c+}} \tag{B.10}$$

The probability that sampling any additional cue results in a positive or negative cue depends on cue reliability and the posterior probability of the resource quality:

$$\Pr(c_+|D) = \sum_{q \in Q} \Pr(c_+|q)\, B(q|D) \tag{B.11}$$

$$\Pr(c_-|D) = \sum_{q \in Q} \Pr(c_-|q)\, B(q|D) \tag{B.12}$$

**1.1.3. Interruption rate.** If an agent chooses to accept a resource, it does not always receive this resource: with a fixed probability $\rho$, its collection of the resource is interrupted. If interruption happens, the resource disappears. The encounter ends here without any further changes to that agent's somatic state.

### 1.2. Extrinsic events

Unless an agent's somatic state equals 0 after a resource encounter, an agent experiences an extrinsic event. Extrinsic events are outside the control of the agent. Just like resources, the extrinsic event quality can range between -20 and +20 and is always a multiple of 0.2. Let $E$ denote the set of all possible extrinsic events:

$$E = \{\, x \,|x = 0.2n, n \in \mathbb{N}_0, -20 \leq x \leq 20\} \tag{B.13}$$

Extrinsic event qualities are independently and identically drawn from a truncated normal distribution with mean $\mu_{extrinsic}$ and standard deviation $\sigma_{extrinsic}$:

$$\Pr(e|\mu_{extrinsic}, \sigma_{extrinsic}) = \frac{\eta(e|\,\mu_{extrinsic}, \sigma_{extrinsic})}{\sum_{f \in E} \eta(f|\,\mu_{extrinsic}, \sigma_{extrinsic})} \quad \text{(B.14)}$$

If an agent is alive after the extrinsic event, it continues to the next cycle.

### 1.3. Fitness

The *immediate* outcome of a cycle is the sum of the outcome of a resource encounter and an extrinsic event. That is, the immediate outcome is the change in somatic state from the start and end of that cycle. Let $O_{immediate}(CY_t|b_t)$ denote the immediate outcome of the $t$'th cycle:

$$O_{immediate}(CY_t|b_t) = b_{t+1} - b_t \quad \text{(B.15)}$$

The *total* expected outcome of the $t$'th cycle is the sum of the immediate outcome, plus the expected outcomes of all future cycles. The expected outcome of future cycle is conditional on the somatic state that an agent will start that cycle with:

$$O_{total}(CY_t|b_t) = O_{immediate}(CY_t|b_t) + O_{total}\big(CY_{t+1}\big|b_t + O_{immediate}(CY_t|b_t)\big) \quad \text{(B.16)}$$

In principle, an agent could go through an infinite number of cycles. However, we assume there is some somatic degradation over time. Due to this decline, the outcomes from cycles that are in the distant future are less valuable than those in the near future. We model discounting as an exponential decrease in expected utility over time. That is, each future cycle is 'worth' a little less than the preceding cycle. If $U(CY_t|b_t)$ is the expected utility (i.e., discounted outcome) the cycle $CY_t$, then:

$$U(CY_t|b_t) = O_{immediate}(CY_t|b_t) + \lambda * O_{total}\big(CY_{t+1}\big|b_t + O_{immediate}(CY_t|b_t)\big) \quad \text{(B.17)}$$

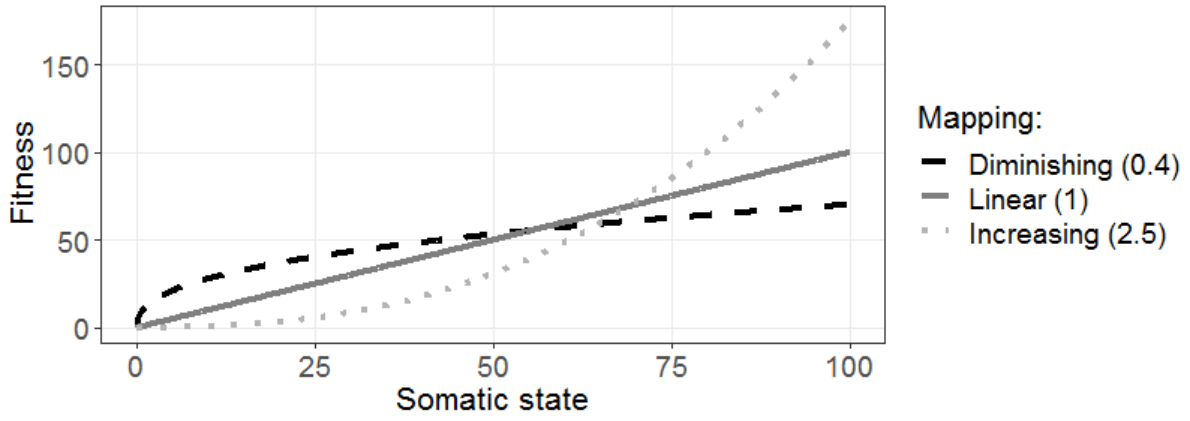


**Fig B.3.** A graphical depiction of the relation between somatic state at the end of life and the fitness associated with that state. We explore three mapping: marginally diminishing returns ($\alpha = 0.2$; solid dark grey line), marginally linear returns ($\alpha = 1$; dashed black line), and marginally increasing returns ($\alpha = 2.5$; dotted light grey line).

Where $\lambda$, *the discount rate*, is a constant between 0 and 1. Lower values of $\lambda$ result in steeper discounting and accordingly in lower expected utility. We set λ to 0.95.

Although an agent goes through an infinite number of cycles as long as its somatic state exceeds zero, there is a theoretical final cycle (see appendix C). After this theoretical last cycle the agent accrues fitness consequences based on its somatic state after that cycle. How this somatic state translate to fitness is determined by the fitness function $\omega(b_{t=\infty})$:

$$\omega(b_{t=\infty}) = \beta {b_{t=\infty}}^{\alpha} \tag{B.18}$$

Where α determines the slope of the curve. We consider three mappings of somatic state to fitness (figure B.3): linear marginal returns (marginal fitness increments are constant as the somatic state increases; $\alpha = 1$); diminishing marginal returns (marginal fitness increments decrease as the somatic state increases; $\alpha = 0.4$); and increasing marginal returns (marginal fitness increments in increase as the somatic state increases; $\alpha = 1$). However, these three functions have different areas under the curve. As a result, there is a higher amount of fitness to be had when the mapping is increasing than in the other two mappings. As the fitness benefits are proportional to sampling costs (that is, sampling is relatively less costly when fitness is more easily accrued), we normalize all three functions with a normalizing constant $\beta$ to ensure that the area under the curve is the same for all mappings:

$$\beta = \frac{\sum_{b \in B} b}{\sum_{b \in B} b^{\alpha}} \tag{B.19}$$

### 1.4. The environment: harshness and unpredictability

We vary the levels of environmental harshness and unpredictability between agents. We explore two notions of harshness: as a (low) mean resource quality $\mu_{resource}$, and as a (low) average extrinsic morbidity $\mu_{extrinsic}$. We also explore two notions of unpredictability: as variance of outcomes given a mean level of harshness, and as a high rate of interruptions. Because we use two interpretations of harshness, this first interpretation of unpredictability applies to both the outcome of the resource encounter and for the extrinsic event. As a result, we vary both the standard deviation of resource quality $\sigma_{resource}$, and of extrinsic events $\sigma_{extrinsic}$. We vary the interruption rate by changing ρ. Let $env$ denote a vector consisting of five state variables that together represent an environment:

$$env = \begin{bmatrix} \mu_{resource} \\ \mu_{extrinsic} \\ \sigma_{resource} \\ \sigma_{extrinsic} \\ \rho \end{bmatrix} \tag{B.20}$$

We systematically explore different combinations of these five state variables, drawing resource and extrinsic event qualities from a discrete normal distribution with a range from -20 to 20. Due to this range restriction, the mean and standard deviation of the normal distribution are not independent. In the most extreme cases, as the means approach 20, the maximum possible standard deviation approaches 0.

This restriction of range also influences the means and standard deviations in less extreme cases. That is, if we draw resources from the distribution specified in equation B.3, the mean and standard deviation of all observed resources is not equal to $\mu_{resource}$

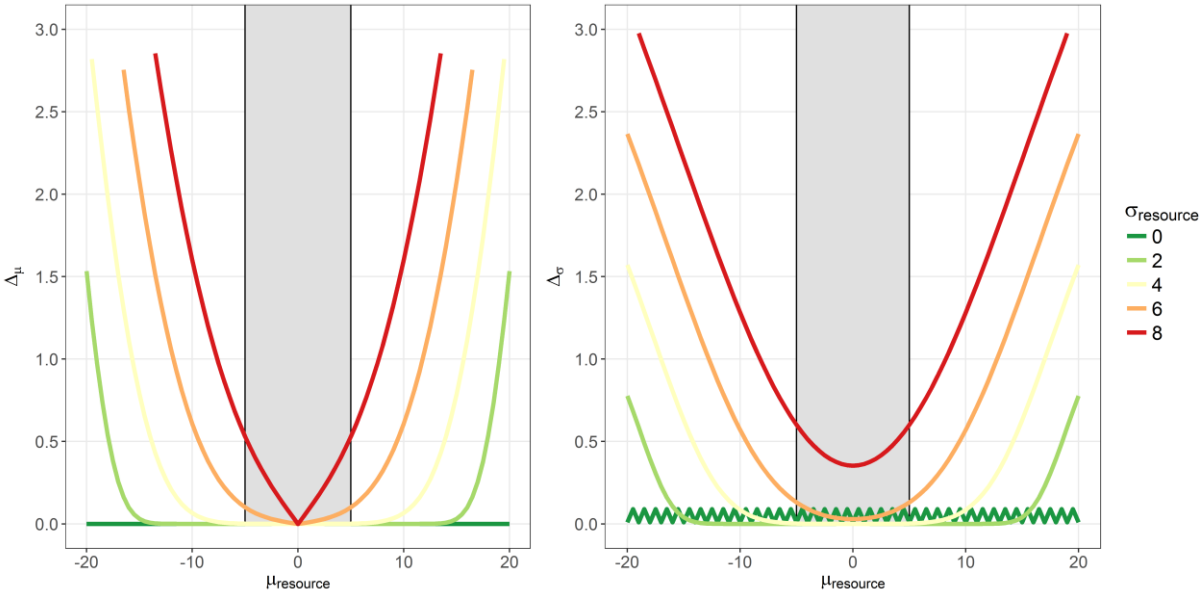


**Figure B.4.**
The absolute differences $\Delta_\mu$ and $\Delta_\sigma$ as a function of $\mu_{resource}$ and $\sigma_{resource}$. We restrict $\mu_{resource}$ to be between -5 and 5 (grey square), and $\sigma_{resource}$ between 0 (dark green) and 8 (orange). Under these ranges, the maximum $\Delta_\mu$ is approximately 0.53 and the maximum $\Delta_\sigma$ is approximately 0.60.

and $\sigma_{resource}$, respectively. Let $\mu_{observed}$ and $\sigma_{observed}$ denote mean and standard deviation of all resources in the environment:

$$\mu_{observed} = \sum_{v \in V} v * Pr(v|\mu_{resource}, \sigma_{resource}) \neq \mu_{resource} \tag{B.21}$$

$$\sigma_{observed} = \sqrt{\sum_{v \in V} ((\mu_{observed} - v)^2 * Pr(v|\mu_{resource}, \sigma_{resource}))} \neq \sigma_{resource} \tag{B.22}$$

Let $\Delta_\mu$ be the absolute difference between $\mu_{resource}$ and $\mu_{observed}$, and let $\Delta_\sigma$ be the absolute difference between $\sigma_{resource}$ and $\sigma_{observed}$:

$$\Delta_\mu = |\mu_{observed} - \mu_{resource}| \tag{B.23}$$

$$\Delta_\sigma = |\sigma_{observed} - \sigma_{resource}| \tag{B.24}$$

Figure B.4 plots $\Delta_\mu$ and $\Delta_\sigma$ as a function of $\mu_{resource}$ and $\sigma_{resource}$. The absolute differences increase rapidly with both extreme means and higher standard deviations. Such large differences are problematic. To minimize these differences, we restrict resource means to the range $[-5, 5]$ with steps of 0.5, and the standard deviations of resource quality to the range $[0.5, 8]$ with steps of 0.5. Under these restrictions the maximum $\Delta_\mu$ is approximately 0.53 and the maximum $\Delta_\sigma$ is approximately 0.60.

Importantly, although above we describe the problem that a range restriction has on resources, the same problem holds for extrinsic events. We therefore restrict the extrinsic means to the range [-5,5] with steps of 5 and standard deviations of extrinsic events to the range [0,8] with steps of 4.

We study environments where the interruption rate ρ is 0, 0.25, or 0.5. We define set of all environments $ENV$ as:

$$ENV = \left\{ \begin{matrix} 0.5n * \mu_{resource} \\ 5n * \mu_{extrinsic} \\ 0.5n * \sigma_{resource} \\ 4n * \sigma_{extrinsic} \\ \rho \end{matrix} \middle| \begin{matrix} -5 \leq \mu_{resource} \leq 5, n \in \mathbb{N}_0 \\ -5 \leq \mu_{extrinsic} \leq 5, n \in \mathbb{N}_0 \\ 0 \leq \sigma_{resource} \leq 8, n \in \mathbb{N}_0 \\ 0 \leq \sigma_{extrinsic} \leq 8, n \in \mathbb{N}_0 \\ \rho \in \{0, 0.1, 0.25, 0.5\} \end{matrix} \right\} \quad \text{(B.25)}$$

These parameters do not change within an environment over cycles. Moreover, the agent lives in a single environment. There is only one agent per environment.