# 1 Description of circular firefly task

The task has two main phases: the **observation phase** and the **action phase**. These correspond to the causal inference and navigation segments of the task, respectively.

**Observation phase.** This phase lasts for an amount of time $t_{obs}$ (= 300 ms for human subjects). The subject begins at position $x_0 = 0$ and moves at a constant (possibly zero) angular speed $w$. The firefly begins at $f_0$, ends at $f_a := f_0 + vt_{obs}$, and moves at a (possibly zero) velocity $v$, i.e.

$$f(t) = f_0 + vt \qquad t \in [0, t_{obs}] \ . \tag{1}$$

From the point of view of the (possibly rotating) observer, the distance $d(t) := f(t) - x(t)$ between the firefly and the observer is

$$d(t) = f_0 + (v - w)t = f_a - (v - w)(t_{obs} - t) \qquad t \in [0, t_{obs}] \ . \tag{2}$$

We assume that the subject continuously receives noisy observations of the relative firefly position and their self-motion during the observation phase. In particular, discretize the time interval $[0, t_{obs}]$ into $N$ time bins, so we have $\Delta t := t_{obs}/N$ and $t_i := i \ \Delta t$. The subject receives $N$ self-motion observations $\{w_1, ..., w_N\}$ and firefly location observations $\{d_1, ..., d_N\}$. Assume these observations are noisy, and in particular that

$$\begin{aligned} d_i &\sim \mathcal{N}(d(t_i), \sigma_f^2/\Delta t) \\ w_i &\sim \mathcal{N}(w, \sigma_w^2/\Delta t) \end{aligned} \tag{3}$$

where $\sigma_f$ is the distance observation noise, and $\sigma_w$ is the self-motion observation noise.

**Action phase.** This phase lasts for an amount of time $T$ (this is 1.5 sec for human subjects). The firefly continues to move (if it was previously moving), but is now invisible, and the subject is under complete control of their movement. In the subject's frame of reference, the firefly begins at a position $f_a$.

The subject's effectively one-dimensional movement is controlled via angular velocity inputs $u(t)$, i.e.,

$$\dot{x} = u(t) \ . \tag{4}$$

The input is assumed to be **bounded**, i.e., $u_t \in [-60 \text{ deg/s}, 60 \text{ deg/s}]$.

In this phase, the firefly's dynamics are assumed to be slightly different than in the observation phase (when it was visible). If its velocity is nonzero, its movement now also involves a constant acceleration component in the same direction as its velocity. The precise formula is

$$f(t) = f_a + v(t - t_{obs}) + \text{sign}(v)\Delta f \left( \frac{t - t_{obs}}{T} \right)^2 \tag{5}$$

where $t \in [t_{obs}, t_{obs} + T]$. The true final position of the firefly is

$$f_T := f(t_{obs} + T) = f_a + vT + \text{sign}(v)\Delta f \ . \tag{6}$$

For human subjects, the acceleration contribution is quite large: $\Delta f = 50$ deg.

## 2 Overview of theory

Our model assumes that well-trained subjects perform each component of the task—causal inference and navigation—'optimally', but possibly with wrong assumptions. This idea has occasionally been called 'rational', as opposed to 'optimal', decision-making.

In particular, we assume subjects construct a belief about firefly motion (position and velocity) and self-motion in accordance with Bayes' rule. The overall belief state is three-dimensional. (There is also an acceleration belief, but this is not determined by within-trial observations.) For convenience, we assume they make a binary decision about the underlying causal structure (i.e., given the available evidence, the firefly either moved or did not) and use their estimates conditioned on that causal structure to determine a steering plan for the navigation part of the task.

We assume subjects navigate 'optimally' in the sense that their steering trajectory minimizes a certain objective. However, we allow this objective to include factors other than reaching the target using the least amount of effort. Hence, trajectories may appear suboptimal from the experimenter's point of view, since the objective includes other terms. The specific objective we assume is

$$-V := \int_0^T \left\{ \frac{\alpha}{2}u^2 + \frac{\kappa^3}{2}\dot{u}^2 - \rho u \dot{f} + \frac{1}{2\beta}(x-f)^2 \right\} e^{-t/\delta}\, dt + \frac{(x_T - f_T)^2}{2} e^{-T/\delta} \,. \tag{7}$$

In words: the objective includes (temporally discounted) moment-to-moment penalties on velocity, acceleration, alignment with the firefly's velocity, and proximity to the firefly, in addition to the experimenter-determined reward for actually reaching the firefly at the end of the trial.

Note that while the endpoint reward is modeled as quadratic, the actual reward profile looks somewhat different. One major difference is that the reward is exactly zero unless the subject is within a certain distance of the firefly (±25 degrees for human subjects). The second difference is that it is pyramid-shaped within that region rather than quadratic.