Using Inverse Planning and Theory of Mind for Social Goal Inference

Sean Tauber (sean.tauber@uci.edu)
Mark Steyvers (mark.steyvers@uci.edu)
Department of Cognitive Sciences, University of California, Irvine
Irvine, CA 92697 USA

Appendix

Figure 3-a and 3-b in the main text show the graphical model for an agent's belief formation and action choice. It is important to note that the "agents" in our experiment did not actually use these models to generate their actions. The agent's were simply animations that displayed many of the possible combinations of orientation and motion relative to the location of the objects in the environment. The graphical models represent our assumptions about the human observers' *theory* about the processes that underlie the agent behavior.

There are $K = 2$ objects, where object $k = 1$ is the agent and object $k = 2$ is the loot; and $L = 6$ discrete locations (two north, two south, and two east).

As depicted in figure 1-c of the main text, there were only $L = 5$ locations that were visible in the experiment (1 north, 2 east, 2 south). We originally assumed that observers would represent the environment with five rooms (because that is what was shown). However, in this case the LS model performed slightly worse than shown in figure 4 (where we used 6 locations) but still better than the other three models. After examining the specific trials that had the greatest discrepancy between the LS model and human responses, we noticed that in many of these trials the agent was moving north, and the humans were responding as if there was a second room to the north separated by a door. Interestingly, it appeared that human observers viewed the maze as being symmetrical in all three directions, and that the northern most room and door were occluded from the display. Based on this finding, we simply changed the representation of the environment to include an additional room to the north with a door between the two northern rooms. Because the objects never actually appeared in this "invisible" room, the main effect of this change was that it changed the agent's prior belief state about the location of the objects.

A.1 Belief Model

The observer's theory about the process by which an agent forms a belief state about the environment is shown in figure 3-a of the main text.

Let s^k be the true location of object k where $p(s^k = x)$ is distributed as a categorical with a uniform prior probability over all locations $x \in \{1..N\}$. Our experiment had $N = 6$ discrete locations (rooms).

$$\alpha_i = \frac{1}{N} = \frac{1}{6}$$
$$s^k \sim categorical(\boldsymbol{\alpha})$$

The agent makes a series of observations where o_i^k is the agent's observation of the presence ($o_i^k = 1$) or absence ($o_i^k = 0$) of object k in location i. If the agent does not observe a location then $o_i^k = -1$. From the agent's perspective, the probability that it observed what it did ($\boldsymbol{o}$) is dependent on whether it orientated towards each location ($\theta_i = 1$) or not ($\theta_i = 0$); whether there is a closed door obstructing each location ($w_i = 1$) or not ($w_i = 0$); and the unobserved true state of the environment ($\boldsymbol{s}$). The only difference between each of the four models (LS, PR, XV, and AK) is in the observer's theory about how the agent's observations are generated based on its orientation, the location of walls, and the true state of the environment. The probability of observing that object k is in location i, denoted ($o_i^k = 1$); observing that object k is not in location i, denoted ($o_i^k = 0$); or not observing location i, denoted ($o_i^k = -1$) is provided for each of the models below.

$$p\left(o_i^k = y \middle| s^k, \boldsymbol{\theta}, \boldsymbol{w}\right) = \begin{cases} 1 \quad , & w_i = 0;\ \theta_i = 1;\ s^k = i\ ;\ y = 1 \\ 0 \quad , & w_i = 1\ ;\ y = 1 \\ 0 \quad , & \theta_i = 0\ ;\ y = 1 \\ 0 \quad , & s^k \neq i\ ;\ y = 1 \\ & \\ 0 \quad , & w_i = 0;\ \theta_i = 1;\ s^k = i\ ;\ y = 0 \\ 0 \quad , & w_i = 1\ ;\ y = 0 \\ 0 \quad , & \theta_i = 0\ ;\ y = 0 \\ 1 \quad , & w_i = 0;\ \theta_i = 1;\ s^k \neq i\ ;\ y = 0 \\ & \\ 1 \quad , & w_i = 1\ ;\ y = -1 \\ 1 \quad , & \theta_i = 0\ ;\ y = -1 \\ 0 \quad , & w_i = 0;\ \theta_i = 1\ ;y = \ -1 \end{cases} \qquad \text{(LS model)}$$

$$p\left(o_i^k = y \middle| s^k, \boldsymbol{\theta}, \boldsymbol{w}\right) = \begin{cases} 1 \quad , & w_i = 0;\ s^k = i\ ;\ y = 1 \\ 0 \quad , & w_i = 1\ ;\ y = 1 \\ 0 \quad , & s^k \neq i\ ;\ y = 1 \\ & \\ 0 \quad , & w_i = 0;\ s^k = i\ ;\ y = 0 \\ 0 \quad , & w_i = 1\ ;\ y = 0 \\ 1 \quad , & w_i = 0;\ s^k \neq i\ ;\ y = 0 \\ & \\ 1 \quad , & w_i = 1\ ;\ y = -1 \\ 0 \quad , & w_i = 0\ ;\ y = -1 \end{cases} \qquad \text{(PR model)}$$

$$p\left(o_i^k = y \middle| s^k, \boldsymbol{\theta}, \boldsymbol{w}\right) = \begin{cases} 1 \quad , & \theta_i = 1;\ s^k = i\ ;\ y = 1 \\ 0 \quad , & \theta_i = 0\ ;\ y = 1 \\ 0 \quad , & s^k \neq i\ ;\ y = 1 \\ & \\ 0 \quad , & \theta_i = 1;\ s^k = i\ ;\ y = 0 \\ 0 \quad , & \theta_i = 0\ ;\ y = 0 \\ 1 \quad , & \theta_i = 1;\ s^k \neq i\ ;\ y = 0 \\ & \\ 0 \quad , & \theta_i = 1\ ;\ y = -1 \\ 1 \quad , & \theta_i = 0\ ;\ y = -1 \end{cases} \qquad \text{(XV model)}$$

$$p\left(o_i^k = y \middle| s^k, \boldsymbol{\theta}, \boldsymbol{w}\right) = \begin{cases} 1 \quad , & s^k = i\ ;\ y = 1 \\ 0 \quad , & s^k \neq i\ ;\ y = 0 \\ 1 \quad , & s^k \neq i\ ;\ y = 0 \\ 0 \quad , & s^k = i\ ;\ y = 1 \\ 0 \quad , & y = -1 \end{cases} \qquad \text{(AK model)}$$

Based on its observations $\boldsymbol{o}^k$, $\boldsymbol{\theta}$, and $\boldsymbol{w}$, the agent is able to infer a posterior probability over s^k using the following formula:

$$p(s^k | \boldsymbol{o}^k, \boldsymbol{\theta}, \boldsymbol{w}) \propto \prod_{i=1..L} p\left(o_i^k \middle| s^k, \boldsymbol{\theta}, \boldsymbol{w}\right) \cdot p(s^k)$$

$$p(s^k = x | \boldsymbol{o}^k, \boldsymbol{\theta}, \boldsymbol{w}) \propto \prod_{i=1..L} p\left(o_i^k \middle| s^k = x, \boldsymbol{\theta}, \boldsymbol{w}\right) \cdot p(s^k = x)$$

We will refer to the posterior distribution $p(s^k | \boldsymbol{o}^k, \boldsymbol{\theta}, \boldsymbol{w})$ as the agent's state of belief about the location of object k. When the observer actually makes an inference about an agent's belief state based on $\boldsymbol{\theta}$, and $\boldsymbol{w}$, we refer to $p(s'^k | \boldsymbol{o}^k, \boldsymbol{\theta}, \boldsymbol{w})$ as the agent's belief state about object k from the observer's perspective.

A.2 Utility

Let t be the agent's type where $t = 1$ is the *cop* type and $t = 2$ is the *robber* type. We define an agent's probability distribution over goals and priorities conditional on their type.

Let g^k be the agent's goal w.r.t. object k where object $k = 1$ is the *agent object* and object $k = 2$ is the *loot object*. Let $g^k = 1$ represent that the agent's goal is to *move towards* object k; and let $g^k = 2$ represent that the agent's goal is to *move away* from object k.

$$P(g^k = 1|t) = \begin{cases} 1 & , \quad k = 1\,;\ t = 1 \\ 1/2 & , \quad k = 2\,;\ t = 1 \\ 0 & , \quad k = 1\,;\ t = 2 \\ 1 & , \quad k = 2\,;\ t = 2 \end{cases}$$

$$P(g^k = 2|t) = \begin{cases} 0 & , \quad k = 1\,;\ t = 1 \\ 1/2 & , \quad k = 2\,;\ t = 1 \\ 1 & , \quad k = 1\,;\ t = 2 \\ 0 & , \quad k = 2\,;\ t = 2 \end{cases}$$

Let c^k be the agent's classification of object k as *primary* ($c^k = 1$), *secondary* ($c^k = 2$), or *irrelevant* ($c^k = 3$).

$$P(c^k = 1|t) = \begin{cases} 1 & , \quad k = 1\,;\ t = 1 \\ 0 & , \quad k = 2\,;\ t = 1 \\ 1 & , \quad k = 1\,;\ t = 2 \\ 0 & , \quad k = 2\,;\ t = 2 \end{cases}$$

$$P(c^k = 2|t) = \begin{cases} 0 & , \quad k = 1\,;\ t = 1 \\ 0 & , \quad k = 2\,;\ t = 1 \\ 0 & , \quad k = 1\,;\ t = 2 \\ 1 & , \quad k = 2\,;\ t = 2 \end{cases}$$

$$P(c^k = 3|t) = \begin{cases} 0 & , \quad k = 1\,;\ t = 1 \\ 1 & , \quad k = 2\,;\ t = 1 \\ 0 & , \quad k = 1\,;\ t = 2 \\ 0 & , \quad k = 2\,;\ t = 2 \end{cases}$$

Each agent type has a utility function that depends on its goals, priorities, and belief state. Irrelevant objects, where ($c^k = 3$), do not affect the utility calculation. Primary objects, where ($c^k = 1$), have the greatest impact on utility. Secondary objects, where ($c^k = 2$), affect the utility insomuch as the goals relating to secondary objects do not conflict with goals relating to primary objects. The goals and priorities are fixed for each of the two agent types. Therefore, we have a utility function for the cop type and a utility function for the robber type. To make the utility function more readable, we replace the location indices $i = 1..L$ with a verbal description of the direction and location number. For example, RIGHT1 is the first room to the right and RIGHT2 is the second room to the right that is behind a closed door.

The utility for a *cop* agent ($t = 1\ , g = \langle 1, \frac{1}{2} \rangle\ , c = \langle 1,3 \rangle$ is,

$$u(a = UP, \boldsymbol{g}, \boldsymbol{c}, \boldsymbol{s}) = P(s^1 = UP1) + P(s^1 = UP2)$$
$$u(a = RIGHT, \boldsymbol{g}, \boldsymbol{c}, \boldsymbol{s}) = P(s^1 = RIGHT1) + P(s^1 = RIGHT2)$$
$$u(a = DOWN, \boldsymbol{g}, \boldsymbol{c}, \boldsymbol{s}) = P(s^1 = DOWN1) + P(s^1 = DOWN2)$$

The utility for a *robber* agent ($t = 2\ , g = \langle 2,1 \rangle\ , c = \langle 1,2 \rangle$ is,

$$u(a = UP, \boldsymbol{g}, \boldsymbol{c}, \boldsymbol{s}) = P(s^1 \neq UP1)\big[1 + P(s^2 = UP1) + P(s^1 \neq UP2)[1 + P(s^2 = UP2)]\big]$$
$$u(a = RIGHT, \boldsymbol{g}, \boldsymbol{c}, \boldsymbol{s}) = P(s^1 \neq RIGHT1)\big[1 + P(s^2 = RIGHT1) + P(s^1 \neq RIGHT2)[1 + P(s^2 = RIGHT2)]\big]$$
$$u(a = DOWN, \boldsymbol{g}, \boldsymbol{c}, \boldsymbol{s}) = P(s^1 \neq DOWN1)\big[1 + P(s^2 = DOWN1) + P(s^1 \neq DOWN2)[1 + P(s^2 = DOWN2)]\big]$$

A.3 Observer Response Model

We define the observer's uncertainty ε about the agent's identity as a function of the binary entropy of the posterior type distribution t where σ is a scale parameter.

$$\boldsymbol{\varepsilon} = \left[-P(t = cop) \cdot ln\big(P(t = cop)\big) - P(t = rob) \cdot ln\big(P(t = rob)\big)\right]^{\sigma} \quad \text{(A.1)}$$

Figure A.1 shows how an observer's uncertainty varies as a function of the posterior type distribution.

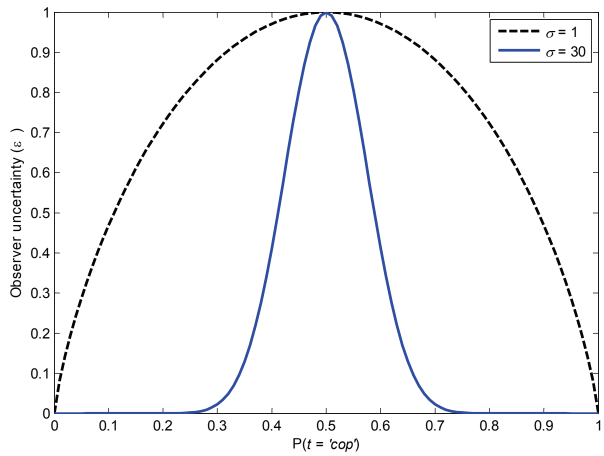

Figure A.1. Observer Uncertainty (ε) as a function of the posterior type distribution (t) and scale parameter (σ).

The observer has maximum uncertainty ($\boldsymbol{\varepsilon} = 1$) when $P(t = cop) = P(t = robber)$ and minimum uncertainty ($\boldsymbol{\varepsilon} = 0$) when $P(t = cop) = 1$ or $P(t = robber) = 1$. We model an observer's response y as:

$$P(y = \text{"}cop\text{"}\ |t) \propto P(t = cop)/(\varepsilon + 1) \quad \text{(A.2a)}$$

$$P(y = \text{"}rob\text{"}\ |t) \propto P(t = rob)/(\varepsilon + 1) \quad \text{(A.2b)}$$

$$P(y = \text{"}don't\ know\text{"}\ |t) \propto \varepsilon/(\varepsilon + 1) \quad \text{(A.2c)}$$