The Gopher's Gambit: Survival Advantages of Artifact-Based Intention Perception

Cynthia Hom, Amani R. Maina-Kilaas, Kevin Ginta,
Cindy Lay, and George D. Montañez

Introduction

- A **designed or real** trap is handmade to be dangerous, while a **random** trap is uniformly sampled from the space of all possible configurations.

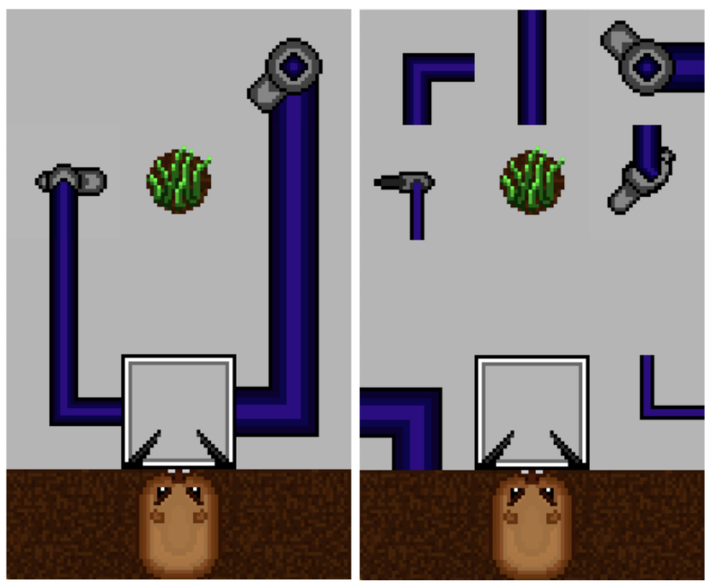


Figure 1: A real trap (left) and a randomly generated trap (right), in our simulated agent world.

- **Baseline** gophers randomly decide whether to enter a trap based on a given probability, while **intention** gophers use the **intention perception algorithm** to assess whether a trap is designed, judging from the **coherence** of its connections, and enters based on that.
- **Cautious** gophers isolate the intention variable by using a "faulty" algorithm that declares traps as designed with the same frequency as the real one, but without connection to the actual trap.

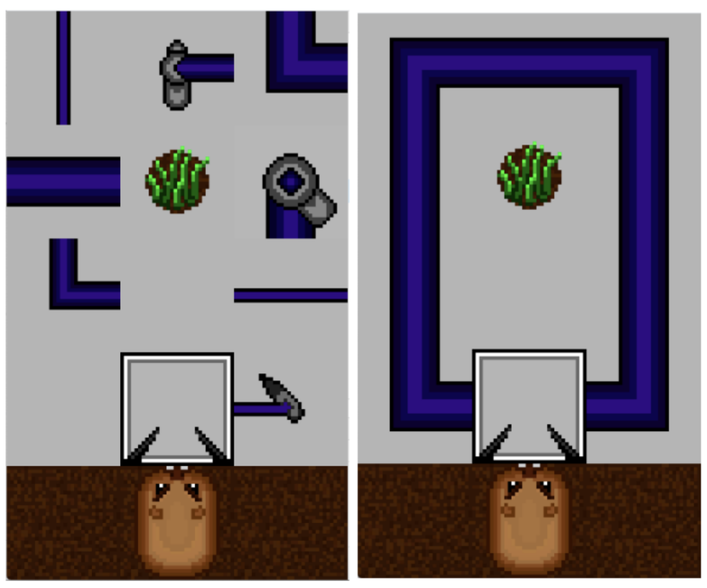


Figure 2: Coherence is correlated with functionality but does not imply it, as exampled by a functional incoherent trap (left) and a nonfunctional coherent trap (right).

Experimental Setup

We vary several probabilities: the baseline gopher's entering a trap, encountering a designed trap, and an **arrow** (the laser-like cell) killing the gopher. The default values are given below. For each set of parameter values we run 10,000 independent trials.

Table 1: Default values for experiment parameters.

Param.	Description	Value
P_e	Prob. of entering trap	0.8
P_r	Prob. of real trap	0.2
$P_{k,w}$	Kill prob. of wide arrow	0.45
$P_{k,n}$	Kill prob. of normal arrow	$\frac{2}{3}P_{k,w}$
$P_{k,s}$	Kill prob. of skinny arrow	$\frac{1}{3}P_{k,w}$
MFI	Maximum Fasting Interval	4

Intention Perception Algorithm

We reject the null hypothesis that a trap is randomly generated at an α level of 0.0001, corresponding to a **surprise value** of 13.29 bits. We calculate the surprise value of a trap configuration with

$$S(x) = -\log_2 \left[|\mathcal{X}|(1 + \ln |\mathcal{X}|) \frac{p(x)}{F_g(x)^{-1}} \right],$$

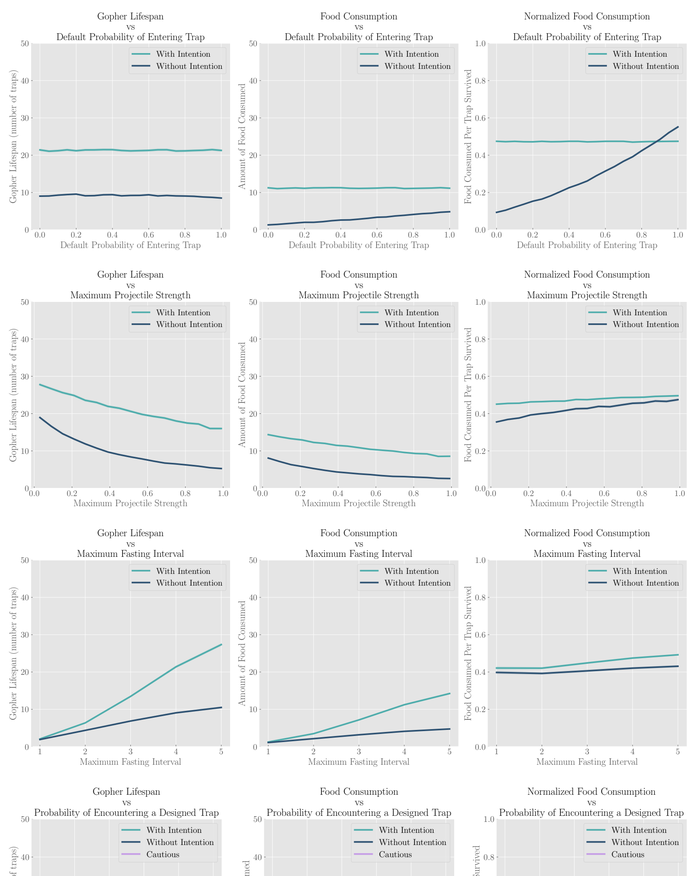
(Montañez, 2018; Hazen et al., 2007), where

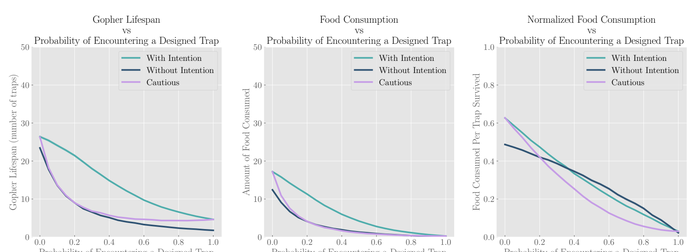
- x is a configuration (i.e., trap)
- $\mathcal{X}$ is the space of all configurations
- $p(x) = 1/|\mathcal{X}|$
- $F_g(x) = M_g(x)/|\mathcal{X}|$
- $M_g(x) = |\{x' \in \mathcal{X} : g(x') \geq g(x)\}|$
- $g(x)$ is the number of coherent connections per nonempty cell of x.

Results

"Signal" in the configurations can be **exploited** through statistical methods, providing **survival advantages.**

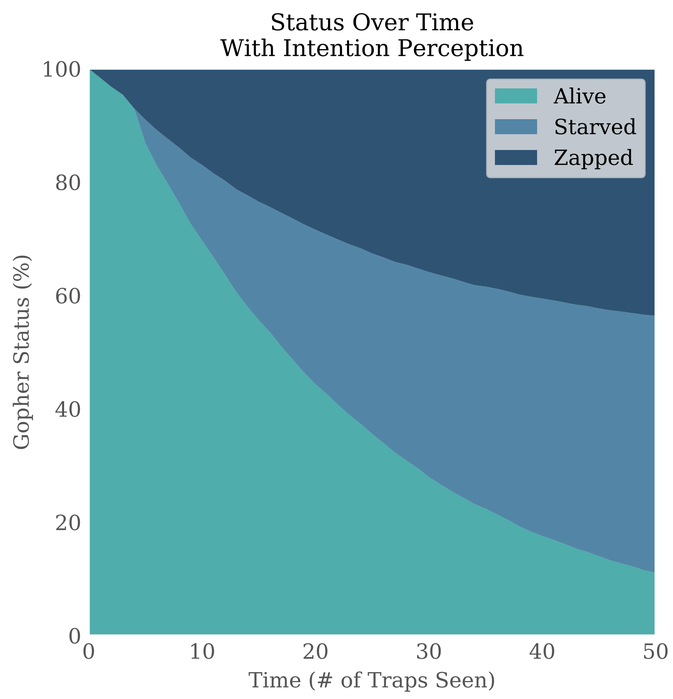
The graphs below show that the intention gophers (light blue) typically have longer lifespans.

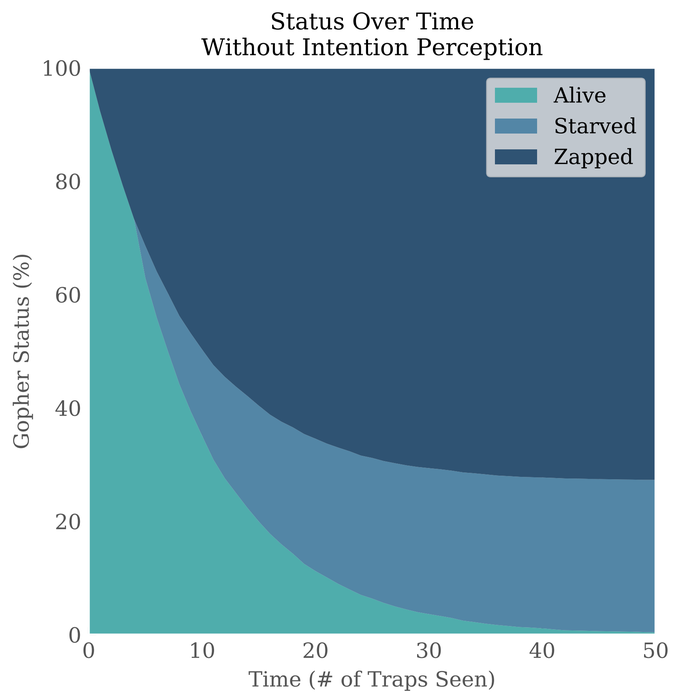




The advantage of intention perception is greatest when **safety is the priority.**

The following plots reveal that intention gophers are more likely to die of starvation than the baseline gophers, but are more likely to remain alive overall.





Conclusion

- **Detection of intentional configurations is possible** (and highly accurate) through statistical analysis of artifacts.
- Knowledge of intention can be **exploited by artificial decision-making systems**.
- Intention perception is **helpful** in a **majority** of tested cases.
- Benefit of intention perception is **greater when prioritizing safety** over food consumption.

Acknowledgments

Special thanks to Jerry Liang, Aditya Khant, Kyle Rong, and Tim Buchheim for assistance in experimental set-up. This research was supported in part by the National Science Foundation under Grant No. 1950885. Any opinions, findings or conclusions expressed are the authors' alone, and do not necessarily reflect the views of the National Science Foundation.