Higher-Order Reasoning under Intent Uncertainty Reinforces the Hobbesian Trap

Otto Kuusela
University of Amsterdam
Amsterdam, The Netherlands
otto.kuusela@outlook.com

Debraj Roy
University of Amsterdam
Amsterdam, The Netherlands
d.roy@uva.nl

ABSTRACT

Civilisations in the universe face the difficulty of communicating and trying to understand others' intentions. Moreover, advanced civilisations could develop weapons to pre-emptively eliminate any civilisations that present a future threat – this is known as the Hobbesian trap. Here, we present a multi-agent simulation model to investigate conditions for such pre-emptive attacks. We design a novel algorithm for solving Interactive Partially Observable Markov Decision Processes (I-POMDPs) with continuous state and observation spaces; it enables civilisations to perform higher-order reasoning. The algorithm builds a nested hierarchy of search forests using Monte Carlo simulations, determining updated beliefs by weighting existing particles. Our experiments reveal interesting insights into the behaviour of rational civilisations under varying levels of reasoning, morality and uncertainty. We find that selfish civilisations always create a war-like universe. Even good, universalist civilisations can initiate pre-emptive attacks if they are uncertain about others' intentions. Finally, our findings have important implications for international peace and security and may explain persistent conflicts and the fragility of ceasefires. Under such conditions a well-coordinated international approach, facilitated by international alliances such as the United Nations, is paramount.

KEYWORDS

Hobbesian trap; pre-emptive action; I-POMDP; higher-order reasoning; civilisations

1 INTRODUCTION

There is at least one civilisation in the universe. What if there are more? This is a possibility that is rooted in credible academic discourse. For example, in [25], the authors estimate the number of communicating extraterrestrial civilisations based on astronomical data and under a range of different assumptions. The existence of other civilisations cannot be ruled out, although it remains a hypothesis in the absence of data.

Our goal is to investigate the interaction of civilisations in the universe. This is an interesting problem to study for two reasons. First, the environment of space does not make finding and making friends easy. Distances are incomprehensibly large and travelling takes time. Even sending messages using the seemingly instant electromagnetic radiation is slow. The second reason is that the civilisations do not know who they are interacting with. The uncertainty about the intentions of the counterpart is one of the defining features of this problem.

The main argument motivating our work is presented in [12]. Civilisations in the universe develop in vastly different cultural contexts. In the absence of a common context in which to interact and learn each others' ways of communication, translating messages sent by other civilisations seems challenging and potentially impossible. Moreover, creating weapons capable of inflicting great damage against other civilisations seems to be easy for sufficiently advanced civilisations. The authors argue that given the lack of communication and the possibility of possession of world-destroying capabilities, civilisations may choose to perform pre-emptive strikes. In other words, they could attack because they fear someone else might attack them. This line of reasoning is called a *Hobbesian trap*. The implication is that contact with extraterrestrial civilisations may be an existential risk for humanity.

We investigate the problem with a computational model. Given the lack of information about how civilisations behave, we ground our work in the assumption that civilisations are *rational*. We propose a novel algorithm for solving I-POMDPs – a framework for rational behaviour – with continuous state and observation spaces in Section 3.2. Finally, we describe the simulation experiments we performed and their results in Section 4. We show that the nature of the universe depends on the morality of civilisations and what they believe about the morality of others. If civilisations are indifferent towards the well-being of other civilisations (and believe others are too), attacks are frequent. Even civilisations that prefer not to destroy others can be motivated to pre-emptively attack if they are unsure about whether a growing civilisation will want to threaten them in the future.

2 BACKGROUND

2.1 Conflict Between Civilisations in Space

Different scenarios of humanity's contact with extraterrestrial civilisations are catalogued in [3]. Beneficial scenarios usually assume successful cooperation and exchange of information. The difficulty in communication, whether related to translation or technology, is acknowledged. In a harmful scenario humanity could be abused for resources or entertainment.

In terms of ethics, civilisations vary on the axis "selfish" versus "universalist" [3]. As opposed to selfish civilisations, universalists value things like life or consciousness of all civilisations, not just themselves. The authors of [9] argue that civilisations are likely to have democratic forms of government since they are more stable than autocracies. Thus, we could also expect them to be peaceful like their counterparts on Earth. However, we argue this peacefulness is not guaranteed to extend to interstellar contacts.

In opposition to [12], another study [13] argues that civilisations in space would be deterred from pre-emptive attacks because of the associated uncertainties. By the time the attacker arrives at the target, the target could be stronger than the attacker. An attack could fail to destroy it entirely and leave behind vindictive scraps of civilisation. Other civilisations could notice the attack and decide that they prefer not having hostile neighbours. Here we investigate the special case where attacks and observations happen without delay (independent of distance) and the target is completely destroyed if it is weaker than the perpetrator. We also assume that civilisations are rational, which means that retaliation is not a reason in itself to attack another civilisation.

2.2 Conflict Between Human Civilisations

Wars between nations are of interest to us, since they are the closest equivalent to interstellar conflict we can study. Wars happen, broadly speaking, for the same reasons smaller conflicts between groups [5, 6] of humans do [21, Ch. 5]. At the heart of many conflicts between human groups is competition over limited resources. Knowing that others may desire to forcefully take one's resources creates *fear*. Fear can lead to a preemptive attack; this mechanism is known as the *Hobbesian trap* [21, Ch. 2] [2]. For example, in the post-World War II U.S. many advocated for a preventive strike against the Soviet Union before they could develop a substantial nuclear weapons capacity [18]. The Hobbesian trap can be modelled using game theory with a conflict game similar to stag hunt [2]. Uncertainty about the opponent's cost of attacking can make attacking a rational choice.

Wars have been shown to have interesting statistical properties [21, Ch. 5] (see also [22] for more visualisations). The starting times of wars follow a Poisson process, and thus times between outbreaks of wars follow an exponential distribution. This means that wars begin randomly and independently at a constant rate. Likewise, the duration of wars follow an exponential distribution. Finally, the numbers of deaths in wars (or *magnitudes* of wars) follow a power-law distribution.

2.3 Modelling Rational Behaviour

Rational behaviour is studied through the paradigm of decision theory, with a central assumption that agents act to maximise their expected utility [19]. If this utility is obtained through a sequence of decisions, the problem is called a *sequential decision problem.* Markov decision processes (MDPs) are a basic model for sequential decision problems. POMDPs generalise MDPs by introducing partial observability of the underlying state. The agent maintains a belief distribution over the state space using observations. [23]

The problem of considering others' decisions in one's own decision-making is studied in *game theory* [19]. Of particular interest here

are *stochastic games* [16]. They, along with their partially observable extension (POSG) [8], capture dynamic situations where decisions cause the game to change. Equilibrium concepts can be used to predict the behaviour of rational agents [10].

While equilibria are appropriate for *describing* or predicting the kind of behaviours one might observe in a multi-agent system, they are less well suited for *prescribing* the best course of action. Acting according to the equilibrium strategy is only optimal when others act according to the same equilibrium. This introduces problems when there are multiple equilibria and no obvious way to know which one other agents are acting according to (if any). Interactive POMDPs (I-POMDPs) provide an alternative approach to modelling multi-agent environments [7]. It is a framework for solving a POSG not from the bird's-eye view of game theory, but from the subjective point of view of the acting agent. It extends a POMDP by including models of other agents in the state space. These models represent the agent's beliefs about others' strategies. Instead of using the concept of an equilibrium to choose its actions, the agent acts optimally with respect to these beliefs.

3 METHODS

In this section we introduce our model and a novel approach to solving I-POMDPs. For readers unfamiliar with the I-POMDP framework, we provide supplementary material in Appendix A.2.

3.1 Modelling Civilisations Using the I-POMDP Framework

Let us create a model of a system of civilisations. The ultimate goal of civilisations is to survive. Each civilisation is characterised by a metric, *technology level*. This variable reflects how advanced the civilisation is when it comes to its capabilities to attack and observe others. The metric ranges from 0 (weak) to 1 (strong) and grows over time. We use *sigmoid growth* which reflects the s-shaped nature of technological growth seen in human societies. It has two parameters, 'speed' g_s and 'takeoff age' g_t, which vary between civilisations. The technology level at age t is given by $\tau(t, g_s, g_t) = 1/(1 + \exp(-g_s(t - g_t)))$. Spatially, civilisations are distributed uniformly and randomly in the two-dimensional unit square. Technology level determines the *radius of influence* of a civilisation. To be able to observe or attack a target, it needs to be within this distance. For technology level τ, the radius is given by $r(\tau) = 0.1 \tan((\pi/2)\tau)$. The radius grows asymptotically as the technology level approaches 1. Civilisations with a technology level higher than approximately 0.96 can influence the entire universe. Civilisations can also attempt to hide their presence from others. They do this by controlling their *technosignatures*, which are signals indicative of intelligent life [20]. We assume it takes the form $v\tau$ where v is the *visibility factor* between 0 and 1 and τ is the technology level of the civilisation. Finally, the initial values of the civilisations' parameters – age, visibility factor, growth speed and takeoff age – independently follow the uniform distributions (discrete, where appropriate) over the intervals $\mathcal{I}_t$, $\mathcal{I}_v$, $\mathcal{I}_{g_s}$ and $\mathcal{I}_{g_t}$, respectively. This joint distribution over a single civilisation's parameters is denoted p_{init}.

In the I-POMDP framework, civilisations use models of other civilisations to predict how they act. These models form a hierarchy

where a *reasoning level* characterises the number of levels of models taken into account by a civilisation. Specifically, we assume that the following models are used. At level 0, civilisations choose actions randomly. The lowest level civilisations in our model reason at level 1. These civilisations model others using level 0 models. In other words, at level 1 civilisations assume they are interacting with opponents that act randomly. At level 2, civilisations think they are playing against a level 1 civilisation. Therefore at level 2 the other agents are modelled with an intentional, or rational, model. Even higher levels are possible, but here we do not go further than level 2. At reasoning level 1 civilisations hold beliefs about the state of the universe. In contrast, at level 2 beliefs are about both the state and the beliefs and frame of the level 1 opponent.

We will now define the components of the I-POMDP ($IS_{i,L}, A, O_i, T, Z_i, R_i, C_i$) of agent i who reasons at level L.

3.1.1 States. Let $N = \{1, \ldots, n\}$ be the set of civilisations. States are tuples $s = (s_1, \ldots, s_n)$, where s_j is the state of civilisation $j \in N$. A civilisation state is a tuple (t, v, g_s, g_t) consisting of the age t (in model time steps), visibility factor v and growth parameters of a civilisation. Locations of civilisations are assumed to be common knowledge and therefore it is not necessary to include them in the state.

3.1.2 Actions. Possible actions for a civilisation include hiding, attacking one of the other civilisations or doing nothing. During one time step of the decision process one randomly chosen civilisation gets to act. Therefore the joint actions $(a_1, \ldots, a_n) \in A$ consist of an actual action by one civilisation while others do a 'no turn' action. The time steps are thought to be relatively short in "universe time". This reflects the hypothesis that civilisations in the real universe observe frequently but act sparsely[1]. Actions therefore take place in a sequence rather than simultaneously.

3.1.3 Observations. Civilisations make noisy observations about others' technosignatures, constrained by their radius of influence. They also observe the results of any attacks on them and attacks initiated by them. Specifically, assume the process is in state $s = (s_1, \ldots, s_n)$ and the previous joint action was a. The resulting observation $o_j \in O_j$ of agent $j \in N$ is a random vector $(\hat{\tau}_1, \ldots, \hat{\tau}_n, r_1, r_2)$ where $r_1, r_2 \in \{-1, 0, 1\}$ indicate the result of an attack by agent j and the result of an attack on agent j, respectively (-1 means that no attack took place). In addition,

$$\hat{\tau}_k = \begin{cases} \tau(s_k) + \Phi_k & d_{jk} = d_{jj} = 0 \\ v\tau(s_k) + \Phi_k & 0 < d_{jk} \leq r(\tau(s_j)) \\ \Upsilon_k & d_{jk} > r(\tau(s_j)) \end{cases}$$

where d_{jk} is the distance between civilisations j and k, $\tau(s_k)$ is the technology level of agent k in state s, Φ_k is a normal random variable (independent of others) with mean 0 and standard deviation σ_{obs} and Υ_k follows a uniform distribution on $[0, 1]$. These distributions define the observation probability function Z_j. The civilisation only receives substantive technosignature observations from civilisations within its radius of influence. It observes its own technology level directly, without the effect of the visibility factor.

3.1.4 Transition function. The transition function T is deterministic. First, at every step the age t of each agent is increased by 1. If a civilisation with a higher technology level attacks another civilisation that is i) within its radius of influence and ii) weaker, the target is destroyed. When this happens, the age t of the target is set to zero and its visibility factor v to one. This reflects the destruction of the civilisation and its capabilities. Taking the hiding action multiplies the visibility factor v by a constant between 0 and 1. We call this constant the visibility multiplier v_m. This multiplicative effect reflects diminishing returns in the attempts to mask one's technology level as observed by others.

3.1.5 Rewards. The reward function R_j determines the expected reward received by civilisation j when an action is taken in a state. If j is destroyed it receives a reward of $r_D = -1$. This is the worst outcome for the civilisation. The cost hiding is a model parameter $r_h \in (-1, 0]$, as is the cost $r_a > -1$ of attacking another civilisation. This latter cost reflects mostly moral considerations as any material and manufacturing costs of a weapon are likely to be negligible in comparison. In all other cases the reward is 0.

3.1.6 Optimality Criterion. Civilisations use the infinite horizon criterion with discounting as their optimality criterion. Assume that civilisation j at level l holds a belief distribution $b_{j,l}$ over the level l interactive states. If the current time is t^*, this means that j attempts to maximise the expected *utility*

$$U(b_{j,l}) = \mathbb{E}\left(\sum_{t=t^*}^{\infty} \gamma^{t-t^*} R_j^{(t)}\right) \tag{1}$$

where $R_j^{(t)}$ is a random variable denoting the reward received by the civilisation on time step t. We denote this criterion C_j. Further explanation can be found in Appendix B.1.

3.2 Novel Algorithm for Solving I-POMDPs

In this section we introduce a new algorithm developed for solving the I-POMDP defined in Section 3.1. The algorithm combines ideas from I-NTMCP [24] (see Appendix B.2.6), LABECOP [11] (Appendix B.2.7) and the Interactive Particle Filter [4] (Appendix B.2.2) to efficiently solve I-POMDPs with continuous state and observation spaces. It constructs a set of forests which form a nested hierarchy.

3.2.1 Structure. The hierarchy of forests reflects the hierarchy of models in the initial belief $b_{i,L}$ of agent i. At the top of the hierarchy at level L is the forest $\mathcal{F}_i(\hat{\theta}_i)$ which represents the decision-making of agent i. If other agents are modelled using intentional models, these are represented by the forests $\{\mathcal{F}_{i,j}(\hat{\theta}_j) \mid j \in N \setminus \{i\}\}$ at level $L-1$. A forest is created for every frame $\hat{\theta}_j$ of agent j that is assigned a positive probability in $b_{i,L}$. A single agent's forests corresponding to its different frames at a given level – denoted $\mathcal{F}_{i,j}$, for example – constitute a *forest group*. This hierarchy continues until level 0, or level 1 if level 0 agents are modelled with a subintentional model.

Each node in forest $\mathcal{F}_{\cdot,j}(\hat{\theta}_j)$ (where $\cdot$ denotes a possibly empty sequence of agents) corresponds to a unique *agent action history* – a sequence of actions – of agent j. If a node corresponds to a time t agent action history, its child nodes correspond to time $t + 1$ histories. A node contains a set of *particles*. A particle p stores a state $s(p)$, a history $h(p)$ of actions taken by all agents (a *joint*

---

[1]This is similar to how modern states spy on their enemies but engage in hostility very rarely.

action history), a tuple of frames $\hat{\theta}(p) = (\hat{\theta}(p)_k)_{k \in N \setminus \{j\}}$ for other agents (assuming they are modelled intentionally), the number of times $n(p, a_j)$ it has been propagated with each action a_j and the utility estimate $u(p, a_j)$ of taking the action a_j and then continuing optimally if p represents the true state. In addition, p remembers its ancestor, which is the particle that was propagated with some joint action to create p. The particles link the forests together: the agent action history $h(p)_k$ and frame $\hat{\theta}(p)_k$ of agent k in particle p point to a specific node in the forest $\mathcal{F}_{\cdot,j,k}(\hat{\theta}(p)_k)$. We can denote this node $\mathcal{F}_{\cdot,j,k}(\hat{\theta}(p)_k)(h(p)_k)$.

3.2.2 Representing and Updating Beliefs. The particles in a node are weighted to represent a belief. A belief is always maintained in the unique root node of the forest $\mathcal{F}_i(\hat{\theta}_i)$. This is agent i's current belief of the state. In addition to the usual properties, the particles in the root nodes of each forest $\mathcal{F}_{\cdot,j}(\hat{\theta}_j)$ contain a belief $\delta(p) = (\delta(p)_k)_{k \in N \setminus \{j\}}$ for other agents. These are weights that can be used to create beliefs in the nodes of other agents p points to in the forests on the level below. In this way the particles in the root nodes represent interactive states.

Updating beliefs is necessary during planning and after a real action is performed and an observation is received. Updating is done according to the Sequential Importance sampling and Resampling (SIR) particle filter approach [1]. Let $b_{j,l}$ be the time-t belief of agent j in the node $\mathcal{F}_{\cdot,j}(\hat{\theta}_j)(h_j)$ and o'_j be the new observation received. First, the belief $b_{j,l}$ is resampled using systematic resampling [17]. The particle p in node $\mathcal{F}_{\cdot,j}(\hat{\theta}_j)(h_j a_j)$ then receives the weight

$$b'_{j,l}(p) \propto b_{j,l}(\text{ancestor}(p)) Z_j(s(p), h(p)(t), o'_j) \qquad (2)$$

where $h(p)(t) = (\ldots, a_j, \ldots)$ refers to the most recent joint action in the history stored in p.

After a real action a_i at time t by the owner i of the I-POMDP, the aforementioned process is used to create a belief in the new unique root node $\mathcal{F}_i(\hat{\theta}_i)(h_i a_i)$ corresponding to time $t + 1$. In addition, it is necessary to determine the beliefs $\delta(p)$ of other agents for each particle p in the time $t + 1$ nodes of each forest $\mathcal{F}_{\cdot,j}(\hat{\theta}_j)$. To do this, first the beliefs of these other agents are initialised using the time-t beliefs stored in the ancestor of p. An observation o'_k is then sampled for each agent $k \in N \setminus \{j\}$ from the distribution $Z_k(s(p), h(p)(t), \cdot)$. Here Z_k refers to the observation probability function in the frame of the corresponding forest. These simulated observations are used in the belief update process and the resulting weights for the relevant time $t + 1$ nodes are stored in $\delta(p)$. After this process is complete, all time t nodes can be removed.

3.2.3 Initialisation. Each forest is initialised with a single root node corresponding to an empty agent action history. A fixed number n_{init} of particles are added to each one. The particles and corresponding beliefs and frames for other agents are sampled according to the initial belief of the I-POMDP.

3.2.4 Planning. Planning is done using Monte Carlo simulations while keeping track of updated beliefs. Each forest group is simulated n_{simul} times. Forests are simulated bottom-up, starting from the lowest-level forest groups. An illustration of planning is shown in Figure 1. A simulation in forest group $\mathcal{F}_{i,j_{L-1},\ldots,j_l} := \mathcal{F}_{\cdot,j}$ begins by determining a particle to start from. In practice this means

weighting the particles in one of the root nodes of one of the forests in the group and sampling the particle according to this initial belief. The belief is determined top-down, guided by the beliefs of the owner. First, a particle (p) is sampled according to the belief in the root node of $\mathcal{F}_i(\hat{\theta}_i)$. The belief of agent j_{L-1} stored in the particle initialises the weights of particles in a root node of a forest in the group $\mathcal{F}_{\cdot,j_{L-1}}$. This process can be repeated until a belief is initialised in a forest in the group $\mathcal{F}_{\cdot,j}$, denoted $\mathcal{F}_{\cdot,j}(\hat{\theta}_j)$. After a starting particle is sampled according to the belief, the beliefs of other agents are initialised using the particle. This creates initial beliefs in the forest groups $\{\mathcal{F}_{\cdot,j,k} \mid k \in N \setminus \{j\}\}$ which are used as models of the other agents' behaviour during the simulation.

Next, a series of forward steps is performed down the chosen tree. The goal of a step is to simulate the rational decision-making of j and the other agents given their beliefs $b_{j,l}$ and $\{b_{k,l-1} \mid k \in N \setminus \{j\}\}$ at some time t. During a step, the current particle p is propagated with an action and new beliefs are formed for the next time step. First, each actor (typically all agents but in our model a randomly chosen agent) chooses an action. For agent j a Monte Carlo Tree Search (MCTS) inspired policy is used. The current estimates of action utilities are calculated. The estimate of a_j is

$$\hat{U}(b_{j,l}, a_j) = \frac{\sum_{\substack{p \in \mathcal{F}_{\cdot,j}(\hat{\theta}_j)(h_j) \\ n(p,a_j) > 0}} b_{j,l}(p) u(p, a_j)}{\sum_{\substack{p \in \mathcal{F}_{\cdot,j}(\hat{\theta}_j)(h_j) \\ n(p,a_j) > 0}} b_{j,l}(p)} \qquad (3)$$

where the summation is over particles in the current node that have been expanded with a_j at least once. Next, define the quantities $N_+(b_{j,l}) = \sum_{p \in \mathcal{F}_{\cdot,j}(\hat{\theta}_j)(h_j), b_{j,l}(p) > 0} n(p)$ and $N_+(b_{j,l}, a_j) = (W(b_{j,l}, a_j)/W(b_{j,l})) N_+(b_{j,l})$ which are "the total number of propagations of positively weighted particles" and "the weighted number of propagations using action a_j", respectively. In the above $n(p)$ is the number of times p has been propagated in total, $W(b_{j,l}) = \sum_{p \in \mathcal{F}_{\cdot,j}(\hat{\theta}_j)(h_j)} b_{j,l}(p) n(p)$, and $W(b_{j,l}, a_j) = \sum_{p \in \mathcal{F}_{\cdot,j}(\hat{\theta}_j)(h_j)} b_{j,l}(p) n(p, a_j)$. Now the action a_j is chosen to maximise

$$\hat{U}(b_{j,l}, a_j) + c_{\text{explr}} \sqrt{\frac{\ln N_+(b_{j,l})}{N_+(b_{j,l}, a_j)}}. \qquad (4)$$

The second term, controlled by c_{explr}, encourages choosing unexplored actions. The actions of other agents are chosen using the already simulated forests in $\{\mathcal{F}_{\cdot,j,k} \mid k \in N \setminus \{j\}\}$. Each action a_k is chosen with a probability proportional to

$$\exp\left(\frac{\hat{U}(b_{k,l-1}, a_k)}{c_{\text{sft}} \cdot \left(1/\sqrt{N_+(b_{k,l-1})}\right)}\right) \qquad (5)$$

which is known as the Boltzmann distribution or the softargmax function. The higher the number $N_+(b_{k,l-1})$ of simulations, the more the probability is concentrated on the best actions. The parameter c_{sft} regulates the strength of this effect. If the other agents are modelled with subintentional models, these models are used instead to determine actions.

The particle p is propagated with the chosen joint action a and a next state s' is sampled from the distribution $T(s(p), a, \cdot)$. Here T is the transition function in the frame $\hat{\theta}_j$. A new particle p' is constructed with state $s(p') = s'$, joint action history $h(p') = h(p)a$,
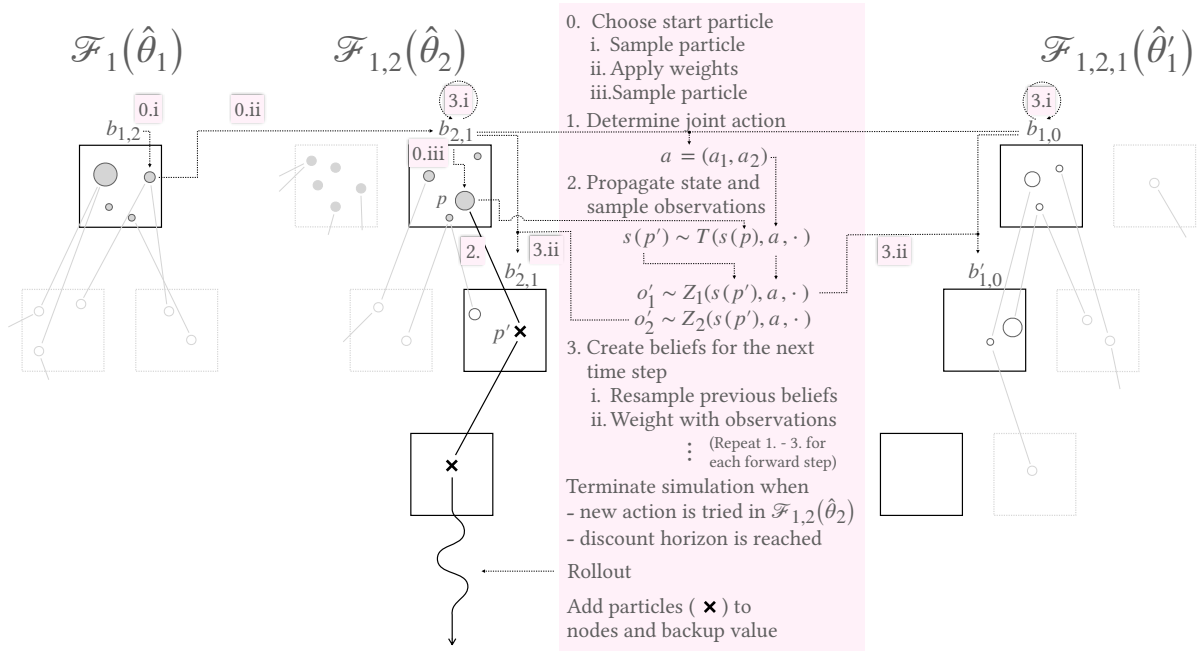
Figure 1: A single simulation in the planning phase. Here the forest $\mathcal{F}_{1,2}(\hat{\theta}_2)$ is being simulated.

and empty propagation information. Observations o'_m are sampled for each agent $m \in N$ according to the observation probability distribution $Z_m(s', a, \cdot)$ from the frame corresponding to the agent's forest. Updated beliefs are generated for the relevant time $t+1$ node in the forest $\mathcal{F}_{\cdot,j}(\hat{\theta}_j)$ and forest groups $\{\mathcal{F}_{\cdot,j,k} \mid k \in N \setminus \{j\}\}$ using the observations. This is the child node of the time t node that corresponds to the chosen action of the forest agent.

Since the resampling step of the belief update process tends to decrease particle diversity, we may optionally add some noise to the state of the new particle before generating the observations. In our model we add unbiased Gaussian noise with standard deviation σ_{g_s} to the growth speed and noise from the uniform discrete distribution $[-\sigma_{g_t}, \sigma_{g_t}]$ to the takeoff age of each civilisation. The perturbed values are then constrained to the ranges $\mathcal{I}_{g_s}$ and $\mathcal{I}_{g_t}$.

A simulation can terminate in two ways. If there is an action a_j for which $W(b_{j,l}, a_j) = 0$, this action is considered unexpanded and is chosen. The current particle p is propagated with a_j and the simulation ends in the resulting particle p'. The second way for a simulation to end is if the discount horizon is reached. This happens when the depth d of the particle p' – the number of time steps traversed forward from the current time – satisfies $\gamma^d < \varepsilon$. The discount horizon $d_{\max}$ is the smallest such number. When a simulation is terminated, a rollout is performed starting from the final particle p'. The state of the final particle is propagated forward with random actions for $d_{\max}$ time steps. The rollout results in a noisy estimate for the utility of taking the action a_j from p (the ancestor of p'). We update $u(p, a_j)$ to equal this estimate.

At the end of the simulation, we traverse the path of nodes taken from the leaf back to the starting root node and add the created particles to the corresponding nodes. The utility estimate $u(p, a_j)$

of taking the chosen action a_j (contained in the joint action a) from particle p is $R_j(s(p), a) + \gamma u(p', \cdot)$, where $u(p', \cdot)$ is the utility estimate from the particle resulting from the propagation of p. In the root node this is combined with previous estimates to be the average of computed estimates thus far. The propagation counter $n(p, a_j)$ is increased by one.

3.2.5 Analysis. We provide a brief analysis of some basic aspects of the algorithm's performance. The number of forests when solving a level L I-POMDP that uses intentional models on all levels and doesn't have uncertainty about frames is $1 + (n-1) + (n-1)(n-1) + \cdots + (n-1)^L = \sum_{k=0}^{L}(n-1)^k = O(n^L)$. This means that planning time is exponential in the reasoning level and polynomial in the number of agents. For the standard two-agent case the number of forests is linear in the reasoning level. Each forest simulation takes $O(n)$ time since the beliefs of each agent need to be updated after each forward step. Note, however, that the simulation time scales superlinearly with the number of simulations done: in later simulations there are more particles to weight during belief update in each encountered node.

4 EXPERIMENTS AND RESULTS

Our goal is to investigate the behaviour of rational civilisations in the universe. We do this by performing the following simulation experiments with our model.

(1) How the morality of civilisations affects the actions taken. We do this by varying the attack reward r_a in the interval $[-0.2, 0.1]$. We measure the proportion of time each action is taken.

(2) The statistical properties of the system when $r_a = 0$. We measure the lengths of streaks of attacks and non-attacks and estimate the underlying distribution.

(3) Investigating pre-emptive attacks in a scenario where a weaker civilisation is potentially surpassing a stronger one in technological capability. We investigate the scenarios $r_a = 0$ ("selfish" civilisations) and $r_a = -0.1$ ("weakly universalist" civilisations).

(4) Investigating the effect of uncertainty about others' intentions. We continue with the scenario of experiment 3, and assume that civilisations are weakly universalist ($r_a = -0.1$). In addition, we assume each agent thinks the other is selfish ($r_a = 0$) with a 50% probability. We want to see if this cost uncertainty elicits pre-emptive attacks.

The parameter values used in the experiments are shown in Table 1 in Appendix C. We compare reasoning levels 1 and 2 to see the effect of higher-order reasoning. The exception is experiment 4, where only reasoning level 2 is appropriate. All experiments use two agents since the interactions in our model are essentially pairwise. The length of simulations is a hundred time steps in experiment 1 and two hundred in experiment 2. In experiments 3 and 4 only the optimal action from the initial belief is computed. The appendix contains model sensitivity analysis (Appendix C) and solver parameter values (Appendix D).

4.1 A Non-negative Attack Cost Leads to a Warring Universe

Here, we investigate the overall behaviour of the system: how often each action is selected by the civilisations. Both civilisations $i \in \{1, 2\}$ begin with an initial belief where i knows its own state s_i exactly but is uninformed about the other's ($\sim i$) state. This means that i thinks $s_{\sim i}$ is distributed according to p_{init}. At reasoning level 2 we assume i thinks $\sim i$ is uninformed by the state (regardless of i's own belief about the state). However, i has no uncertainty about the observational capabilities, rewards and transition function of $\sim i$ and as such uses only the frame $\hat{\theta}_{\sim i,1} = (A, O_{\sim i}, T, Z_{\sim i}, R_{\sim i}, C_{\sim i})$.

The result of varying the attack reward is shown in Figure 2 (left). It shows the proportions of attack actions over hundred time step simulations. Hiding actions (not shown) are taken fairly consistently around 10% of the time, independent of r_a. The rest of the actions are "no action" actions. We observe a transition at $r_a = 0$. Above this value, attacks make up roughly half of all the actions. The reason for this proportion is that typically simulations in this reward range have a strong civilisation that constantly attacks the weaker civilisation. When the weak civilisation acts (roughly half of the time), any attack attempts are counted as "no action" since the stronger civilisation is not within its radius of influence. Below the $r_a = 0$ cut-off there are very few attacks. As we will see in experiment 3 and 4, this does not mean that there are no attacks at all; instead, they are performed only when they are deemed necessary.

We examine the case of selfish civilisations, in other words $r_a = 0$, in experiment 2. Here attacks are costless and but also benefit-less. We measure attack streaks (a series of consecutive attack actions) and peaceful streaks (a series of non-attacks). We performed eight
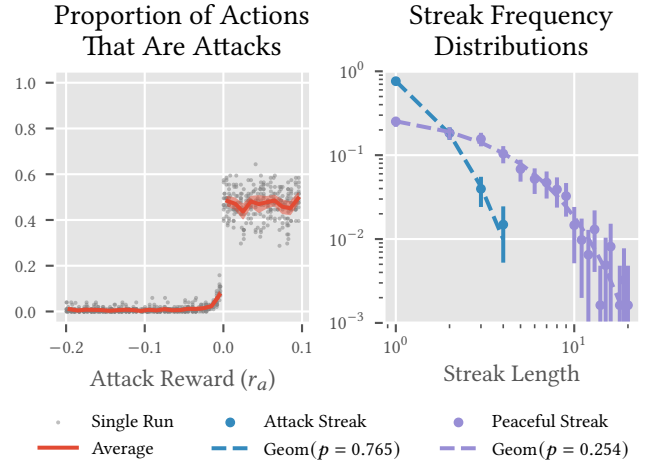

Figure 2: Left: proportion of actions that are attacks. The red shaded region shows a 95% confidence interval for the average. Right: distributions of lengths of attack and non-attack (peace) streaks. The dashed lines show the probability mass functions of fitted geometric distributions.

simulations at both reasoning level 1 and 2. Figure 2 (right) illustrates the distribution of attack streaks observed. In both cases, a geometric distribution is a reasonably good model for the data. Based the estimated values of p, and remembering that typically a strong agent dominates a weaker one in simulations, we can hypothesise the following. At $r_a = 0$ our model is approximately statistically equivalent to a situation where one of the civilisations is chosen at random to act. When selected, the stronger civilisation attacks with a probability of approximately 50%. The weaker civilisation may choose any action. This means that attacks happen with a probability of approximately 25%, which is close to the estimated 25.5%. We find that frequent attacks mean the weak civilisation is, indeed, very weak and thus the strong civilisation cannot plan far enough into future to recognise the potential benefit of attacking the weak civilisation preemptively.

4.2 Pre-Emptive Attacks Depend on How the Other Civilisation is Modelled

Here, we attempt to investigate preemptive attacks: the fear that another civilisation will surpass a civilisation in technological capacity and could therefore be a risk in the future *(surpass scenario)*. Let us denote the two civilisations in the scenario W ("weak") and S ("strong"). We generate an initial belief where the technology level of W is lower than that of S. In addition, we assign a probability to the belief that W surpasses S. At reasoning level 1, with probability p^1_{surpass} W will surpass S within t_{surpass} time steps. At reasoning level 2, the civilisation (either W or S) believes that W will surpass with probability p^2_{surpass} and believes that the other civilisation believes that the surpass will happen with probability p^1_{surpass}. For example, if we consider the situation from the point of view of W and $p^2_{\text{surpass}} = p^1_{\text{surpass}} = 1$, then W is certain it will surpass S ($p^2_{\text{surpass}} = 1$) and it believes S is certain W will surpass
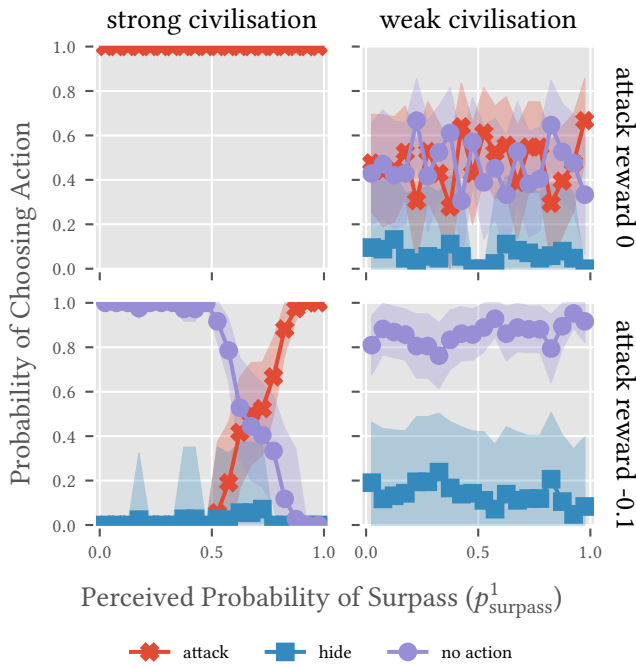
Figure 3: Optimal actions for a level 1 civilisation in the surpass scenario. Columns correspond to strong and weak civilisations. Rows show the actions when attack reward is 0 (civilisations are "selfish") and -0.1 (civilisations are "weakly universalist"). The vertical axis shows the proportion of times that different actions were chosen.
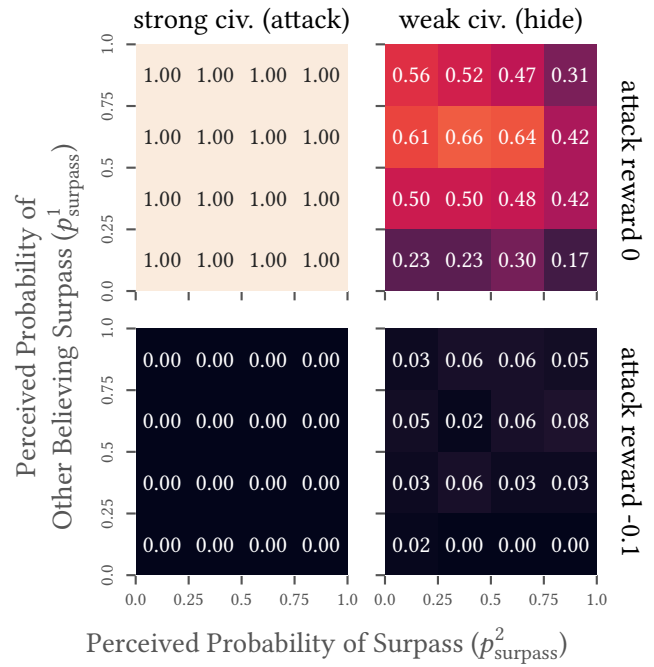

Figure 4: Probability of choosing select actions for a level 2 civilisation in the surpass scenario. For a strong civilisation (left) attacking is shown: for a weak civilisation (right) hiding is shown. Rows show the probabilities when attack reward is 0 (civilisations are selfish) and -0.1 (civilisations are weakly universalist).

it ($p^1_{\text{surpass}} = 1$). We still assume there is no uncertainty about the frame of the other civilisation at reasoning level 2.

We solve the optimal action from this belief for each agent. This is performed for different combinations of p^1_{surpass}, p^2_{surpass} and t_{surpass}. We investigate two scenarios: civilisations are selfish ($r_a = 0$) and weakly universalist ($r_a = -0.1$). The result is shown in Figure 3 for reasoning level 1. When attacking is free (top row), S always attacks W before it gets the chance to grow stronger. W does not have good choices and mostly chooses between attacking (unsuccessfully) and doing nothing. Hiding is avoided because its only benefit is in changing the beliefs of S, and at level 1 W doesn't model these beliefs. When attacking is costly (bottom row) S employs a more sophisticated approach: it is more likely to attack the more certain it is W will surpass it. W abstains from attacking to avoid the cost, but otherwise mostly chooses to do nothing.

Figure 4 shows the corresponding result for civilisations at reasoning level 2, where civilisations reason about each others' beliefs. When civilisations are selfish, S still always attacks, independent of what it believes W believes (top left graph). The top right graph, corresponding to W in a selfish universe, is more intriguing. We have to be careful in the interpretation here: the confidence intervals (not shown) are in the range $\pm[0.14, 0.23]$ for all cells. Since the usefulness of hiding comes from changing beliefs, we would expect the frequency of hiding to depend only on the belief W has about the beliefs of S. There appears to be a region in the middle part of

the belief space where hiding is most useful when W thinks this is where S is the most persuadable: hiding is more likely to take the beliefs of S to a part of the belief space where it does not think W will surpass it. The reason this explanation is incomplete is because, as we saw in Figure 3 (top left), S always attacks at level 1 when there is a positive probability that it will be surpassed. A possible explanation, then, is that the hiding action makes this probability zero.

The bottom left graph corresponds to the frequency of S attacking W in a universe where both are weakly universalist. The graph shows that S never attacks. This is in contrast to Figure 3 (bottom left), where the probability of attacking depends on the beliefs of S. This difference can be explained by considering the models of W that S uses at reasoning levels 1 and 2. At level 1, W is modelled as a level 0 civilisation, i.e. as taking random actions. Such a civilisation is a threat post-surpass, since it can randomly choose to attack S. At reasoning level 2 W is modelled as a level 1 civilisation. Such a civilisation has nothing to gain from attacking S, even if it surpasses S: if that was the case, the now-weaker S could not threaten it.

Finally, in a weakly universalist universe W almost never hides (bottom right graph). Such a civilisation is playing against a level 1 S who attacks when it is reasonably certain it will be surpassed (Figure 3, bottom left). A possible explanation for why hiding is not optimal here is that while the upper part of this graph is where
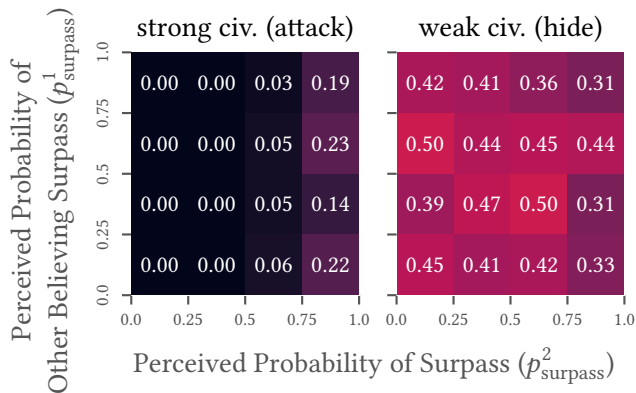
Figure 5: Probability of choosing select actions for a level 2 civilisation in the surpass scenario. For a strong civilisation (left) attacking is shown: for a weak civilisation (right) hiding is shown. Both civilisations are weakly universalist: they have an attack reward $r_a = -0.1$. However, both think that the other's attack reward is -0.1 (weakly universalist) with a 50% probability and 0 (selfish) with a 50% probability.

hiding could help deter attacks, it is also the part where S is the most certain about its belief and thus least persuadable.

4.3 Cost Uncertainty Promotes Pre-Emptive attacks

Here we investigate the case where each civilisation assigns a 50% probability to the other being selfish ($r_a = 0$) and an equal probability to them being weakly universalist ($r_a = -0.1$). In the I-POMDP framework this corresponds to the level 2 civilisation assigning equal probabilities to the two possible frames of the other civilisation. These frames differ only in the reward function; otherwise they are the same frame we have been using thus far. In our algorithm we assign these frames randomly to the interactive state samples representing the level 2 beliefs of the civilisation. The assignment is independent of the environment state in the interactive state sample or the associated beliefs ascribed to the other agent.

The results are shown in Figure 5. It is best compared to the bottom row of Figure 4. The left graph, which visualises how often S would choose to attack in the surpass scenario, shows an interesting difference. Namely, if S is relatively certain that W will surpass it (large $p_{surpass}^2$), then it sometimes engages in a pre-emptive attack. The frequency with which this happens is up to around 23%, although the confidence intervals (not shown) are large at slightly under ±25%. While the exact frequencies are uncertain, the result is qualitatively different to Figure 4: when there is no uncertainty about the reward, S never attacks. This difference is because a selfish W that surpasses S can be a danger to S, prompting a pre-emptive strike.

The fraction of times W chooses to hide (right graph) show a similarly notable difference. W chooses to hide between 30-50% of the time. While there is no discernible pattern in the action choices as $p_{surpass}^2$ and $p_{surpass}^1$ vary (especially considering the error margin of ±17 − 21%), this is again a qualitative difference to Figure 4 (bottom right). It appears that hiding becomes more

rational once there is uncertainty about whether S will launch a pre-emptive attack. Essentially, W is uncertain about whether it is interacting with a selfish level 1 civilisation (Figure 3, top left) which always attacks or with a weakly universalist level 1 civilisation (Figure 3, bottom left) which only attacks when it is relatively sure a surpass will happen. Again, we would expect the frequency with which hiding actions are taken to only depend on the beliefs W has about the beliefs of S. It is not possible to reliably assess this hypothesis due to large uncertainties in the data.

5 DISCUSSION

Our experiments revealed insights into the behaviour of rational civilisations. If attacking is free, attacks are frequent. If attacking is costly (due to moral considerations, for example), the number of attacks significantly decreases. Even so, attacks are still possible. Specifically, a strong civilisation may engage in a pre-emptive attack if it believes another will surpass it in strength and if it is uncertain about the intentions of this growing civilisation. We found this happens in two cases: when the strong civilisation doesn't model the other civilisation as a rational agent, and when it does but is uncertain about how costly attacking is to the growing civilisation. Out of all the experiments, this last scenario is perhaps the most realistic. Based on this, we conclude that if our universe resembles the the model built here, it seems possible for civilisations to fall into the Hobbesian trap and attack out of fear. The findings illuminate the Hobbesian nature of our modern society, where pre-emptive war for control of resources from geopolitically weaker nations is often observed. Further, Hobbesian traps may explain the fragility of ceasefires, especially if the costs of attacks are low and under uncertainty about the intent of the adversary. Under such conditions, a well-coordinated international approach, facilitated by international alliances such as the United Nations, is paramount to preserve peace.

The applicability of these results to our universe hinges on the validity of the assumptions built into our model. Perhaps the most important is the assumption that civilisations in the universe act rationally. Another founding assumption in our work is that civilisations cannot communicate with each other. While perfect translation of an alien language seems extremely challenging, it may be possible for civilisations to communicate simpler messages such as "I am not a threat". Finally, one of the biggest assumptions is that civilisations know of the existence and location of other civilisations in the universe. This is clearly not a valid assumption: discovering other intelligent life is incredibly challenging. In the real universe hiding from stronger, yet-unknown civilisations and seeking hidden, unpredictable civilisations may be important features that are not captured by our model.

Future work should concentrate on extending the results to more than two agents. As discussed above, work should also commence on investigating the applicability of open agent models – models where agents are uncertain about the set of interacting agents – to our work.

6 SUPPLEMENTARY MATERIAL

The code [14] and appendices [15] are available.

7 ETHICS STATEMENT

This ethics statement reflects our commitment to conducting modelling and simulation activities ethically, responsibly, and in the best interest of all stakeholders.

REFERENCES

[1] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing* 50, 2 (2002), 174–188. https://doi.org/10.1109/78.978374

[2] Sandeep Baliga and Tomas Sjöström. 2012. The Hobbesian Trap. In *The Oxford Handbook of the Economics of Peace and Conflict*, Michelle R. Garfinkel and Stergios Skaperdas (Eds.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195392777.001.0001

[3] Seth D. Baum, Jacob D. Haqq-Misra, and Shawn D. Domagal-Goldman. 2011. Would contact with extraterrestrials benefit or harm humanity? A scenario analysis. *Acta Astronautica* 68, 11-12 (June 2011), 2114–2129. https://doi.org/10.1016/j.actaastro.2010.10.012

[4] Prashant Doshi and Piotr J. Gmytrasiewicz. 2009. Monte Carlo Sampling Methods for Approximating Interactive POMDPs. *J. Artif. Int. Res.* 34, 1 (March 2009), 297–337.

[5] Carsten K. W. De Dreu and Zegni Triki. 2022. Intergroup conflict: origins, dynamics and consequences across taxa. *Philosophical Transactions of the Royal Society B: Biological Sciences* 377, 1851 (April 2022). https://doi.org/10.1098/rstb.2021.0134

[6] Donelson R. Forsyth. 2010. *Group Dynamics* (5 ed.). Wadsworth, Cengage Learning.

[7] Piotr J. Gmytrasiewicz and Prashant Doshi. 2005. A Framework for Sequential Planning in Multi-Agent Settings. *J. Artif. Int. Res.* 24, 1 (July 2005), 49–79.

[8] Eric A. Hansen, Daniel S. Bernstein, and Shlomo Zilberstein. 2004. Dynamic Programming for Partially Observable Stochastic Games. In *Proceedings of the 19th National Conference on Artifical Intelligence* (San Jose, California) *(AAAI'04)*. AAAI Press, 709–715.

[9] Albert A Harrison. 2000. The relative stability of belligerent and peaceful societies: implications for SETI. *Acta Astronautica* 46, 10 (2000), 707–712. https://doi.org/10.1016/S0094-5765(00)00035-7

[10] João P. Hespanha and Maria Prandini. 2001. Nash equilibria in partial-information games on Markov chains. In *Proceedings of the 40th IEEE Conference on Decision and Control (Cat. No.01CH37228)*, Vol. 3. 2102–2107. https://doi.org/10.1109/CDC.2001.980562

[11] Marcus Hoerger and Hanna Kurniawati. 2021. An On-Line POMDP Solver for Continuous Observation Spaces. In *2021 IEEE International Conference on Robotics and Automation (ICRA)* (Xi'an, China). IEEE Press, 7643–7649. https://doi.org/10.1109/ICRA48506.2021.9560943

[12] Karim Jebari and Niklas Olsson-Yaouzis. 2018. A Game of Stars: Active SETI, radical translation and the Hobbesian trap. *Futures* 101 (2018), 46–54. https://doi.org/10.1016/j.futures.2018.06.007

[13] Janne M. Korhonen. 2013. MAD with aliens? Interstellar deterrence and its implications. *Acta Astronautica* 86 (2013), 201–210. https://doi.org/10.1016/j.actaastro.2013.01.016

[14] Otto Kuusela. 2024. civilisation-conflict. https://github.com/kuuotto/civilisation-conflict/releases/tag/aamas2024

[15] Otto Kuusela and Debraj Roy. 2024. Higher-Order Reasoning under Intent Uncertainty Reinforces the Hobbesian Trap (Appendix). https://doi.org/10.5281/zenodo.10631669

[16] Kevin Leyton-Brown and Yoav Shoham. 2008. *Essentials of Game Theory: A Concise Multidisciplinary Introduction.* Springer Cham.

[17] Tiancheng Li, Miodrag Bolic, and Petar M. Djuric. 2015. Resampling Methods for Particle Filtering: Classification, implementation, and strategies. *IEEE Signal Processing Magazine* 32, 3 (May 2015), 70–86. https://doi.org/10.1109/MSP.2014.2330626

[18] Karl P. Mueller, Jasen J. Castillo, Forrest E. Morgan, Negeen Pegahi, and Brian Rosen. 2006. *Striking First: Preemptive and Preventive Attack in U.S. National Security Policy* (1 ed.). RAND Corporation. http://www.jstor.org/stable/10.7249/mg403af

[19] Roger B. Myerson. 1997. *Game Theory: Analysis of Conflict.* Harvard University Press.

[20] NASA Technosignatures Workshop Participants. 2019. NASA and the Search for Technosignatures: A Report from the NASA Technosignatures Workshop. arXiv:1812.08681 [astro-ph.IM]

[21] Steven Pinker. 2012. *The Better Angels of Our Nature: A History of Violence and Humanity.* Penguin Books.

[22] Max Roser, Joe Hasell, Bastian Herre, and Bobbie Macdonald. 2016. War and Peace. *Our World in Data* (2016). https://ourworldindata.org/war-and-peace

[23] Stuart J. Russell and Peter Norvig. 2010. *Artificial Intelligence: A Modern Approach* (3 ed.). Pearson.

[24] Jonathon Schwartz, Ruijia Zhou, and Hanna Kurniawati. 2022. Online Planning for Interactive-POMDPs using Nested Monte Carlo Tree Search. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 8770–8777. https://doi.org/10.1109/IROS47612.2022.9981713

[25] Tom Westby and Christopher J. Conselice. 2020. The Astrobiological Copernican Weak and Strong Limits for Intelligent Life. *The Astrophysical Journal* 896, 1 (2020), 58. https://doi.org/10.3847/1538-4357/ab8225