Goal Recognition for Rational and Irrational Agents

Peta Masters
RMIT University
Melbourne, Australia
peta.masters@rmit.edu.au

Sebastian Sardina
RMIT University
Melbourne, Australia
sebastian.sardina@rmit.edu.au

ABSTRACT

Contemporary cost-based goal-recognition assumes rationality: that observed behaviour is more or less optimal. Probabilistic systems, however, generate probability distributions on the basis of suboptimality. We show that, when an observed agent is only slightly irrational (suboptimal), state-of-the-art systems produce counter-intuitive results. We present a definition of rationality appropriate to situations where the ground truth is unknown, define a rationality measure (RM) that quantifies an agent's expected degree of suboptimality, and present a novel self-modulating probability distribution formula for goal recognition. Our formula recognises suboptimality and adjusts its level of confidence accordingly, thereby handling irrationality—and rationality—in an intuitive, principled manner.

ACM Reference Format:
Peta Masters and Sebastian Sardina. 2019. Goal Recognition for Rational and Irrational Agents. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 9 pages.

1 INTRODUCTION

In this paper, we examine—and improve upon—the response of cost-based goal recognition (GR) when confronted by an agent that *seems* less rational than was previously assumed, whether because they are deliberately deceptive, actually irrational (e.g., mad or drunk) or simply operating under an unanticipated cost model.[1]

GR is the problem of determining an agent's intent by observing their behaviour. Traditionally, GR can be categorised into three types: *keyhole* recognition, whereby the observed agent behaves just as if they were not being watched; *intended*, in which they actively attempt to reveal their goal; and *deceptive* or adversarial, where the agent deliberately misleads or obfuscates [2].

Many contemporary cost-based GR accounts implicitly assume keyhole recognition and that the observed behaviour is rational and honest [12, 14, 15, etc.]. On discovering that it is not (i.e., the behaviour increasingly deviates from any optimal plan), one would expect such systems to become more "agnostic", that is, less confident in their predictions until, in the extreme case—confronted by wildly suboptimal or intentionally deceptive behaviour [7]—they judge all goals to be "equally probable".

Contemporary GR models, however, do not achieve this. Technically, there are three issues. First, it is not immediately obvious how rationality should be defined in a domain where the ground truth is unknown. Second, even when observed behaviour is palpably erratic (i.e., unquestionably irrational), systems persist in assessing it as if it were not. Thus, in the absence of meta-reasoning to "notice" that the observed agent (for example) favours now this goal, now that goal, existing GR systems can return apparently conclusive results (e.g., Goal A with probability 0.75) even though two or three observations later the result is reversed (e.g., Goal B with probability 0.8). Third, the probability distribution formulas that we examine—though highly-respected and based on sound principles—actually exhibit unexpectedly anomalous outcomes.

In this paper, we analyse the probability distribution formulas used in support of two state-of-the-art GR frameworks: that of Ramirez and Geffner [12], which focuses on STRIPS-style task-planning; and Vered et al. [17], which focuses on motion-planning. The Ramirez and Geffner framework, in particular, has been used as the basis for many other models and extensions [e.g., 3, 6, 13, 14]. It is significant, therefore, to discover (as we do here) that—when faced with suboptimality in some clear cases—it predicts the most likely goal with *greater* confidence as the agent's perceived behaviour becomes more *ir*rational (clearly a counter-intuitive result); while, in a similar situation, the Vered et al. formula returns distributions that increasingly predict as most probable whichever goal is *most expensive* to reach.

Nevertheless, we build on the insights underlying those formulas to arrive at a more robust model. If the agent's behaviour is rational (as GR systems typically expect), our formula returns a result at the "limit" of the Ramirez and Geffner formula. If the behaviour is irrational, our formula recognises it as such and, while maintaining the overall rankings of possible goals, presents them with decreasing levels of confidence. Using our account, the more irrational the agent's behaviour, the *lower* the confidence and the more probabilities even out to resemble a uniform distribution. Generally speaking, the new formula is "self-modulating" and requires neither meta-processing nor intervention. Moreover, it is a parsimonious elaboration of existing accounts and conceptually simple.

The ability to modulate the outcomes of a GR system confers practical benefits: (i) by preventing it from naively jumping to unwarranted conclusions; (ii) by avoiding oscillation between incompatible decisions; or (iii) if confidence drops *below* a given level, by flagging the possibility of deceptive intent.

In what follows, we first offer an abstract definition of probabilistic GR. We present a definition of rationality for use in domains where the ground truth is unknown, then examine how two state-of-the-art GR systems handle apparent irrationality (i.e., suboptimal plans) and demonstrate anomalies in their outcomes. We define a measure that can be used to calculate an agent's expected degree of rationality based on their past behaviour, then, using that measure, we present and analyse our self-modulating formula. We conclude with a brief review of related work.

[1]Rationality in this paper is synonymous with cost-sensitivity. We say an optimal plan is fully rational and define classes of rationality based on the conditions under which one plan's cost exceeds another's. We use "they" as a singular gender-neutral pronoun.

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

2 BACKGROUND

In this section, we set out the cost-based or "plan recognition as planning" approach to probabilistic GR.

GR involves determining an agent's goal by observing its behaviour. *Probabilistic* GR is the problem of determining the *most likely* goal from a set of possible goals. Whereas traditional GR infers the goals by comparing observations with plans stored in a plan library, *cost*-based GR (in common with other contemporary approaches, such as the use of landmarks [10]) dispenses with that overhead. Instead, it takes a domain theory as part of the input to the GR problem (along with initial state, candidate goals and a partial plan or trajectory) and generates a solution using an off-the-shelf planner. Importantly, the intuition is that a rational (i.e., cost-sensitive) agent is most likely pursuing a rational (i.e., optimal) plan. By comparing the *observed* plan so-far with the *optimal* plan for each goal, a probability distribution can be generated.

Cost-based GR has been studied with respect to various domains, including STRIPS-style task-planning [12, 14], continuous motion-planning [17] and graph-based path-planning [6]. For technical convenience, we express the various accounts as instances of a general GR problem, whose components have different meanings and structures depending on the interpretation given to the domain.

Definition 2.1. A ***cost-based probabilistic GR problem*** is a tuple $\mathcal{P} = \langle \mathcal{D}, O, s_0, G, \vec{o}, Prob \rangle$ where: $\mathcal{D}$ is a model of the GR domain (which defines states, transitions between states and their cost); O is the set of observable elements in $\mathcal{D}$; s_0 is the initial state; G is the set of candidate goals; $\vec{o}$ is a partial sequence of observations drawn from O; and $Prob$ is the prior probability distribution across G.

In this paper, we are particularly concerned with the approaches taken by Ramirez and Geffner [12] and Vered et al. [17] (denoted as ***R&G*** and ***V&K***, respectively). In each case, $\mathcal{D}$ is static and deterministic, the initial state is fully observable and the observation sequence is partial, in that it may not account for the agent's complete behaviour, and not noisy. In R&G, which applies GR to task-planning, $\mathcal{D}$ represents a STRIPS-like domain of fluents and actions $\langle F, A \rangle$ where each action has an associated precondition, add and delete list, all subsets of F. Grounding $\mathcal{P}$ to R&G, O is the set of actions, s_0 is a state under $\mathcal{D}$, G is a set such that each $g \in G$ is a conjunction of literals and $\vec{o}$ is a sequence of actions. In V&K, which is concerned primarily with continuous motion-planning, the domain is conceived as a multi-dimensional Euclidean space, $\mathcal{D} \subseteq \mathbb{R}^n, n \geq 2$, representing two- or three- dimensional map or real-world locations (and capable of representing additional continuous dimensions such as pose, velocity or colour). Grounding $\mathcal{P}$ to V&K, O is a (potentially infinite) set of points and trajectories (or transitions) through $\mathcal{D}$, s_0 is a state (a subset of $\mathcal{D}$), G is a set of such states, $\vec{o}$ is a sequence of points and trajectories, obtained as the range of a time function $f : \vec{t} \mapsto O$, where $\vec{t}$ is a sequence of time intervals during which the world has been observed.

Generally, a ***plan*** π in $\mathcal{D}$ is a sequence of elements that imply changes to the underlying domain, transforming it from one state to another. In R&G, those elements are actions, which are costed, and the ***cost of a plan*** is the sum of its action costs. In V&K, a plan is a trajectory costed by reference to the Euclidean distance metric, so the cost of a plan is, effectively, its length. A plan $\pi = e_1, \ldots, e_m$ is said to ***satisfy observations*** $\vec{o} = o_1, \cdots, o_n$, if there exists a monotonic function $f : \{1, \cdots, n\} \mapsto \{1, \ldots, m\}$ such that $e_{f(i)} = o_i$ for all $i \in \{1, \cdots, n\}$. That is, the order of elements (in both the plan and the observation sequence) is preserved.[2] The ***optimal (lowest) cost*** of a plan from s_0 to a goal $g \in G$ that satisfies observations is denoted by $optc(s_0, \vec{o}, g)$ and when $\vec{o} = \emptyset$ we just write $optc(s_0, g)$.[3]

The ***solution*** to $\mathcal{P}$ is a posterior probability distribution which prefers goals whose plans best satisfy the observations $\vec{o}$, that is, plans that satisfy the observations *at least additional cost* when compared with the cost of an *optimal* plan for the same goal. Intuitively, the more closely observations conform to the optimal plan for a goal g, the more likely it is that goal g is being pursued. Models vary, however, about the preferred method of performing the comparison.

The R&G solution to $\mathcal{P}$ derives from Bayes' Rule and makes two assumptions: that the probability of a plan is inversely proportional to its cost and that probabilities for multiple plans for the same goal can be said to be dominated by the highest of those probabilities. The first assumption is central to their model and is encapsulated in the notion of ***cost difference***, that is, the difference between the cheapest plan for a goal $g \in G$, given the observed actions already taken $optc(s_0, \vec{o}, g)$, and the cheapest plan that could have reached the goal, if one or more of those observed actions had *not* occurred. Following Masters and Sardina [6], we denote the cost of this negative plan as $optc^{\neg}(s_0, \vec{o}, g)$. Formally, cost difference is a function $costdif : S \times O^* \times S \mapsto \mathbb{R}$ defined, when applied to $\mathcal{P}$, as:

$$costdif(s_0, \vec{o}, g) = optc(s_0, \vec{o}, g) - optc^{\neg}(s_0, \vec{o}, g).^4 \quad (1)$$

Ramirez and Geffner's key intuition is that any solution to $\mathcal{P}$ should have the property that *the lower the cost difference for a particular goal, the higher its probability.*

Concretely, this is achieved by plugging the cost difference parameters into a Boltzmann equation, as follows:[5]

$$P_{RG}(G|\vec{o}) = \alpha \cdot \frac{1}{1 + e^{-\beta(optc^{\neg}(s_0,\vec{o},g) - optc(s_0,\vec{o},g))}}, \quad (2)$$

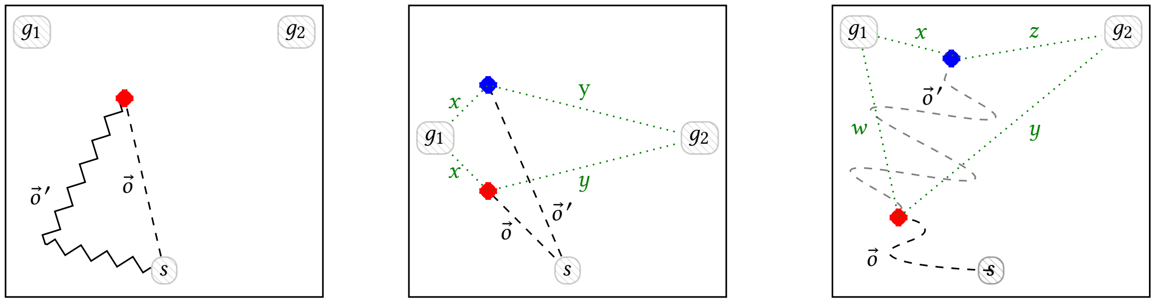
where α is a normalising constant and β is the rate parameter, that is, a positive constant that modifies the default distribution in such a way that, as it approaches zero, the distribution flattens out. It is this rate parameter that we exploit later on. Seeking a similar outcome in the continuous domain and concerned particularly with *online* motion-planning (i.e., observations are assumed to be revealed incrementally and GR is an iterative, rather than a one-off, process), Vered et al. [17] take a different approach. Whereas R&G derives its formula from Bayes' Rule, V&K appeals to empirical evidence to support the use of a ratio between "optimal cost" and "optimal cost through the observations". The authors characterise

[2] Observations may not always be so closely related to plans; the mapping holds, however, for the domains under consideration here. A comparable result may be achieved for other domains by modifying the function. For example, Sohrabi et al. ([2016]) treats observations as observable fluents (not actions), which map into the states of a plan's execution trace rather than mapping to actions in the plan itself.

[3] Though we recognise the abuse of notation, we use $o \in \vec{o}$ to indicate that o occurs somewhere in $\vec{o}$ and $\emptyset$ to represent an empty sequence.

[4] In previous work [9], we have shown that the term $optc^{\neg}(s_0, \vec{o}, g)$ can be replaced in formula (5) by $optc(s_0, g)$ in all but one corner case. Here, however, to avoid conflating orthogonal issues, we use the formula precisely as given by Ramirez and Geffner [12].

[5] We render formula 2 as in [11] but note, the denominator could be re-written as $1 + e^{\beta \, costdif(s_0,\vec{o},g)}$. Though omitted for legibility, for all probability distribution formulas in this paper, values may be multiplied by priors before normalisation.



(a) Equivalence: observations end up in the same state.

(b) Goal-cost equivalence: cost from end-state to each goal is the same.

(c) Relative equivalence: progress towards each goal is the same ($w - x = y - z$).

Figure 1: Irrationality in a GR domain. Whatever the goal, observations $\vec{o}\,'$ are uniformly less rational than $\vec{o}$. Note (a) is a special case of (b) which is a special case of (c). Though the diagrams imply path-planning, the notions are generally applicable.

their probability formula as an heuristic and, indeed, the supporting evidence demonstrates that it was the best performing of three competing heuristics for intent recognition when compared with human performance [1].

Concretely, the probability distribution across G in V&K is based on a simple ratio between the costs of (a) an optimal plan and (b) an optimal plan that satisfies the observations:

$$P_{VK}(G|\vec{o}\,) = \alpha \cdot \frac{optc(s_0, g)}{optc(s_0, \vec{o}, g)}. \tag{3}$$

In the next section, we examine the performance of formulas (2) and (3) more closely. In doing so, we will distinguish between ***scores***, that is, the likelihoods calculated *before* normalisation (which may sum to any value) and ***probability values***, that is, the normalised results (which, after multiplication by a constant α, sum to 1).

3 THE RATIONALITY ASSUMPTION

In this section, we consider how rationality applies in a GR domain. We analyse the above approaches, identify their limitations with respect to the rationality assumption and expose some anomalies. Finally we suggest a mechanism, inspired by V&K, by which we can measure an agent's expected degree of rationality based on their past performance.

3.1 What is Irrational in a GR Domain?

The intuition underlying cost-based GR rests on the assumption of rationality: the more closely an agent is following an optimal plan for A, the more likely it is that A is its intended objective. In the context of a GR problem, however, rationality is not as clear-cut as it would be in a classical planning or a path-planning problem. Normally, we would say that the less rational plan is the one that is more expensive with respect to the real goal but, in a GR scenario, the ground truth is unknown. The fact that observations seem to suggest a plan that is irrational (suboptimal) with respect to any one particular goal actually tells us very little. When an agent pursues a goal, we expect observations to reflect a more-or-less optimal plan for that goal. It stands to reason that the closer the agent is to achieving one goal, the more suboptimal its actions are likely to become with respect to all the others.

Consider, for example, a cooking domain. There are three candidate goals: A, fried eggs, B, boiled potatoes or C, chicken soup. Now, an agent observed peeling potatoes and filling a pan with water is on a more or less optimal path for goal B but an increasingly suboptimal one for goals A and C. Is the plan rational or irrational? truthful or deceptive? Without knowing the real goal, it seems we cannot answer the question. Consider an alternative sequence of observations, however, where the agent is observed heating the oven: a meaningful action in itself, but irrelevant to all three goals! In this case, without needing to know the ground truth, we can confidently describe the observed behaviour as irrational: whatever the goal, it is suboptimal. Similarly, if an agent behaves at one moment as if attempting to achieve goal A, at the next goal B, and so on (e.g., by getting out the frying pan, peeling potatoes and opening a can of soup), now their actions have again become suboptimal with respect to all goals. This is the behaviour that we are interested in: behaviour that indisputably betrays irrationality even though the ground truth is unknown.[6]

Definition 3.1. Observation sequence $\vec{o}\,' \in O^*$ is ***strictly less rational*** than $\vec{o} \in O^*$ iff for all $g \in G$, $optc(s_0, \vec{o}\,', g) > optc(s_0, \vec{o}, g)$.

In words, if one plan (via $\vec{o}\,'$) costs more than another plan (via $\vec{o}$) *no matter which goal is being pursued*, then it is strictly less rational to select the more expensive plan.

For the purpose of demonstrating the limitations of existing systems, we now extend the above definition to describe the special case where an observation sequence is not only strictly less rational than another observation sequence but less rational *by the same degree for all goals*, as follows.

Definition 3.2. Observation sequence $\vec{o}\,' \in O^*$ is ***uniformly less rational*** than $\vec{o} \in O^*$ iff:

(1) $\vec{o}\,'$ is strictly less rational than $\vec{o}$; and
(2) for all $g_1, g_2 \in G$, $optc(s_0, \vec{o}\,', g_1) - optc(s_0, \vec{o}, g_1) = optc(s_0, \vec{o}\,', g_2) - optc(s_0, \vec{o}, g_2)$.

Under this definition, we distinguish three distinct classes of uniformly irrational behaviour, illustrated at Figure 1, as follows.

(a) **Equivalence.** Given two observation sequences such that, no matter which goal is being pursued, the optimal plan that satisfies them ends up in the same state, the one that costs more

[6] This is related to epistemic notions of belief and knowledge under 'possible world' semantics [4]: we believe a plan is irrational if it is irrational *in every possible world* (i.e., for every goal).

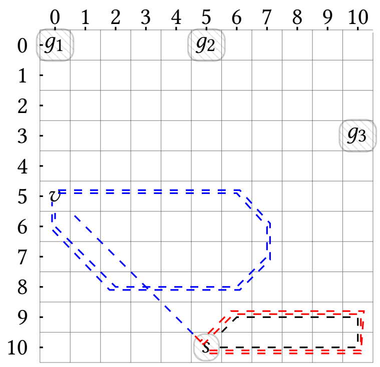


Figure 2: Loopy paths.

Table 1: Probabilities for Loopy Paths.

	R&G ($\beta = 1$)			**R&G ($\beta = 0.1$)**			**V&K**		
	g_1	g_2	g_3	g_1	g_2	g_3	g_1	g_2	g_3
s_1	-	-	-	-	-	-	*0.3333*	0.3333	0.3333
s_2	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	***0.3610***	0.3280	0.3110
s_3	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	***0.3700***	0.3259	0.3042
v_1	***0.9693***	0.0304	0.0003	***0.4200***	0.3343	0.2458	**0.4517**	0.3194	0.2289
v_2	***0.9842***	0.0157	0.0001	***0.4656***	0.3238	0.2106	**0.4162**	0.3224	0.2615
v_3	***0.9842******	0.0157*	0.0001*	***0.4789***	0.3196	0.2014	**0.4060**	0.3223	0.2717
v_{10}	***0.9842******	0.0157*	0.0001*	***0.4820***	0.3186	0.1994	**0.3929**	0.3216	0.2855
	Non-sigmoidal distribution (does not change)								
v_{1-k}	***0.9842***	0.0157	0.0001	***0.4820***	0.3186	0.1994			

Probabilities on each of multiple visits to s and v (see Figure 2). Winners are highlighted. Anomalies are italicised. (* Values have changed by tiny amounts, concealed by rounding.)

is uniformly less rational. This conforms to our usual understanding depicted by $\vec{o}\,'$ and $\vec{o}$ in Figure 1a : a suboptimal plan between states is less rational than an optimal plan between the same two states.

(b) **Goal-cost equivalence.** Given two observation sequences such that, depending on which goal is being pursued, the optimal plan that satisfies them ends up in *different* states from which the cost to each goal is nevertheless the same, the one that costs more is uniformly less rational than the other. This situation is depicted by $\vec{o}\,'$ in Figure 1b.

(c) **Relative equivalence.** Given two observation sequences such that the optimal plan that satisfies them ends up in different states but the cost difference between plans is the same (for all goals), the one that costs more is uniformly less rational than the other. This literal rewording of the definition most obviously arises if the suboptimal path zigzags, as in Figure 1c, advancing on all goals without favouring any one in particular; but the definition subsumes examples (a) and (b), above.

3.2 Analysis of the Rationality Assumption

Definitions 3.1 and 3.2 in hand, we now consider the navigational scenarios depicted in Figures 2 and 3. An agent is observed in a gridworld domain with three goals $G = \{g_1, g_2, g_3\}$. In terms of a GR problem $\mathcal{P}$, the starting location, goals G and observables O are all possible grid locations and transitions are costed in terms of notional edges between adjacent cells, such that horizontal and vertical transitions cost 1, diagonal transitions cost $\sqrt{2}$.

In Figure 2, the bottom right (red) path shows an agent first observed at its initial state s, then moving in two loops; that is, instead of progressing it returns to the cell at s each time. The other (blue) path depicts an agent setting off on an apparently optimal path towards goal g_1. Having reached location v, however, instead of continuing on, it loops twice on location v. In turn, Figure 3 shows the agent again on an apparently optimal path towards g_1 via cell v but this time it veers off to w, then takes an increasingly irrational route via x, y and z (final destination unknown).

Table 1 shows probability values for each goal on each visit to s and v, with an additional result given for the case where the loop returning to v repeats 10 times. There are three main columns: two for the R&G model, each with different β values (the lower β results in a flatter distribution overall); and one for the V&K model (which has no rate parameter). We note the following results.

First, excepting the corner case when the *only* observation is the initial state (shown here as s_1),[7] one can observe that P_{RG} as per Equation (2) evaluates probabilities for paths that repeatedly return to the initial state as equal for all goals (i.e., equivalent to priors). If paths track first to v, however, and then loop, *whichever goal was most probable on the first visit becomes more probable at each subsequent visit!*

Now, this anomaly is not just an issue with "looping" behaviours per se, but with the more general case of *uniformly less rational* observed behaviour, which includes meaningless noise, such as looping (a variation on example (a) above), as a special case. Indeed, our first key result shows that given two observation sequences $\vec{o}$ and $\vec{o}\,'$, goal recognition under Equation (2) becomes *more* confident under the uniformly *less* rational observation sequence $\vec{o}\,'$.

Theorem 3.3. *Let $\vec{o}, \vec{o}\,' \in O^*$ be observation sequences s.t. $\vec{o}\,'$ is uniformly less rational than $\vec{o}$ and let $\hat{g} \in G$ be s.t. $P_{RG}(\hat{g} \mid \vec{o}) > P_{RG}(g \mid \vec{o})$, for all $g \in G \backslash \{\hat{g}\}$ (i.e., goal $\hat{g}$ is the best explanation under the more rational observations $\vec{o}$). Then, $P_{RG}(\hat{g} \mid \vec{o}\,') > P_{RG}(\hat{g} \mid \vec{o})$.*

Proof. The effect is a by-product of normalising scores generated using the Boltzmann distribution.

(1) Let $a = e^{costdif(s_0, \vec{o}, g)}$. Without loss of generality, take $\beta = 1$. Now formula (2) can be rewritten as:

$$P_{RG}(G \mid \vec{o}) = \alpha \cdot \frac{1}{1+a};$$

and we introduce an alternative non-sigmoidal distribution:

$$P_X(G \mid \vec{o}) = \alpha \cdot \frac{1}{a}.$$

(2) Considering the scores (i.e., likelihood of each goal prior to normalisation), clearly, $\frac{1}{1+a} < \frac{1}{a}$. Furthermore, $\lim_{a \to \infty} \frac{1}{1+a} \div \frac{1}{a} = 1$. Thus, as a approaches infinity, P_{RG} converges towards—though it never reaches—P_X.

[7]Owing to the negative reasoning in cost difference equation (1), if there exists a goal g such that *every* path to g satisfies the observations, cost difference may evaluate to $-\infty$ yielding an undefined normalised score [6].

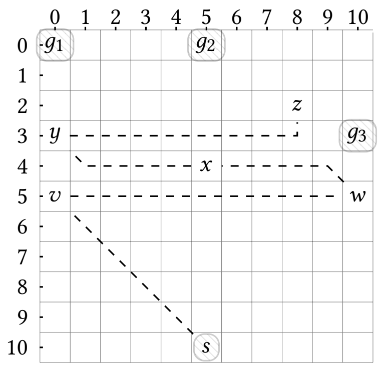


Figure 3: A Zigzagging Path.

Table 2: Probabilities on a Zigzagging Path.

	R&G (β = 1)			**R&G (β = 0.1)**			**V&K**		
	g_1	g_2	g_3	g_1	g_2	g_3	g_1	g_2	g_3
v	**0.9694**	0.0303	0.0003	**0.4200**	0.3342	0.2458	**0.4517**	0.3194	0.2289
w	0.0008	0.0156	**0.9835**	0.2484	0.3264	**0.4351**	0.3176	0.3176	**0.3647**
x	0.3368	**0.6050**	0.0581	0.3436	**0.3609**	0.2955	***0.3708***	0.3380	0.2911
y	**0.9951**	0.0049	4.5e-05	**0.4943**	0.3072	0.1984	***0.4233***	0.3174	0.2593
z	0.0200	0.3734	**0.6066**	0.2694	0.3568	**0.3737**	***0.3562***	0.3318	0.3120

Probabilities as calculated at points v - z (see Figure 3). Anomalies italicised, as previously.

(3) Now, $\frac{1}{a}$ is precisely (by definition) inversely proportional to a, whereas $\frac{1}{1+a}$ is not. Calculating the difference between them, we get:

$$\frac{1}{a} - \frac{1}{a+1} = \frac{(a+1)-a}{a(a+1)} = \frac{1}{a^2+a};$$

which, proportionally, is a decrease of:

$$\frac{1}{a^2+a} \div \frac{1}{a} = \frac{1}{a^2+a} \times \frac{a}{1} = \frac{1}{a+1}.$$

So the proportional decrease is greatest when a is lowest.

(4) Recall that $a = e^{costdif(s_i,\vec{o},g)}$ and $P_{RG}(\hat{g} \mid \vec{o}) > P_{RG}(g \mid \vec{o})$ was given. Therefore, $\frac{1}{e^{costdif(s_i,\vec{o},\hat{g})}+1} > \frac{1}{e^{costdif(s_i,\vec{o},g)}+1}$ for all $g \in G \setminus \{\hat{g}\}$. So, from 3, the proportional decrease $\frac{1}{a+1}$ is more for $\hat{g}$ than for other goals. Therefore, $P_{RG}(\hat{g} \mid \vec{o}) < P_X(\hat{g} \mid \vec{o})$.

(5) Since $\vec{o}\,'$ is less rational than $\vec{o}$ (as given), $costdif(s_0, \vec{o}\,', g) > costdif(s_0, \vec{o}, g)$, for all $g \in G$, and hence $e^{costdif(s_0,\vec{o}\,',g)} > e^{costdif(s_0,\vec{o},g)}$ (i.e., a increases). From points 2 and 4, we get $P_{RG}(\hat{g} \mid \vec{o}) < P_{RG}(\hat{g} \mid \vec{o}\,') < P_X(\hat{g} \mid \vec{o}\,')$. □

The above anomaly occurs whenever cost difference increases uniformly for all goals (i.e., whenever one observation sequence is uniformly less rational than another). The alternative observation sequence does not need to be wildly suboptimal; even the slightest suboptimality generates the same anomalous result.

Note that, although the above result at first appears to contradict R&G's principle—that lower cost difference should result in higher probability—actually, it does not. That principle applies to the total distribution *across goals* with respect to *one* GR problem, whereas Theorem 3.3 examines the situation across *two different problems* (because we have substituted for observations $\vec{o}$ the less rational $\vec{o}\,'$). So, we do not challenge the R&G principle. Nevertheless, Theorem 3.3 states that: when we change to a uniformly less rational observation sequence (a new GR problem), although the relative order across goals is maintained (the R&G principle applies) their *specific probability values* change in counter-intuitive ways.

The discrepancy exposed by Theorem 3.3 arises out of the use of a sigmoidal equation which is then normalised. As soon as we substitute a non-sigmoidal equivalent, the problem is resolved (as proved above and illustrated in the final row of Table 1).

Thus, replacing the R&G formula with its non-sigmoidal counterpart corrects an inconsistency. As a sidenote, we observe that, while this may make only a minor difference to the probability values produced under R&G, such a correction has important ramifications elsewhere. In previous work on GR in path-planning, we proposed an observation-free (or "single-observation") cost difference formula [6, 8]. Using the Boltzmann equation to arrive at the scores, it was necessary to add a large constant to results in order to generate probability values close to parity with those based on the more conventional cost difference formula. Using the Boltzmann, our results maintained only the same *rankings* as Ramirez and Geffner's model; substituting its non-sigmoidal counterpart, however, the probability values returned by their formula would be *identical* to R&G (excepting one corner case described there).[8]

Turning now to V&K, whereas R&G maintains rankings but increases the probability of the most probable goal, faced with irrational paths, this distribution has the effect that the *furthest* goal inevitably becomes more probable than any of the other goals (see highlighted anomalies in Tables 1 and 2).

Theorem 3.4. *Let $\vec{o}, \vec{o}\,' \in O^*$ be observation sequences such that $\vec{o}\,'$ is uniformly less rational than $\vec{o}$ and let $g_1, g_2 \in G$ be such that $optc(s_0, g_1) > optc(s_0, g_2)$ (i.e., the optimal cost of achieving g_1 is greater than that of achieving g_2). Then, there exists a c such that when $optc(s_0, \vec{o}\,', g_1) - optc(s_0, \vec{o}, g_1) \geq c$, $P_{VK}(g_1 \mid \vec{o}\,') > P_{VK}(g_2 \mid \vec{o}\,')$.*

Proof. (Sketch) The numerators for $P_{VK}(\cdot)$ formula (3) are different but constant, based on the optimal cost to each goal. Under the uniformly less rational observations, by Definition 3.2, the cost of the denominator increases equally for all goals. Thus, all scores decrease, but the score with the largest numerator decreases most slowly. Since $optc(s_0, g_1) > optc(s_0, g_2)$, the score for g_1 has the largest numerator, and the proposition follows. □

One impact of Theorem 3.4 is that the cost difference principle is not maintained: lower cost difference does *not* imply higher probability. Furthermore, the fact that, as the size of the denominator increases, it is the *most distant/expensive* goal that begins to be favoured, is again anomalous: the underlying intuition for

[8] We attributed that problem to the use of negative exponentials and loss of precision [6, p.757]. It appears, however, that we were experiencing the effect described in Theorem 3.3 but in reverse. Whereas here we *add* to the cost difference (because paths are suboptimal), there we *deducted* from cost difference (because the single-observation formula dispenses with large portions of the observed path) so the sigmoidal effect described here skewed results in the opposite direction.

cost-based GR is that "cheaper is better" yet here the goal's score increases precisely because its attainment costs more.

Observe also that V&K always returns a comparatively flat distribution: even on the first visit to v on an optimal path to g_1, it yields $P(g_1) < 0.5$. In a practical application, where the user might be waiting for the probability of a goal to exceed some threshold before triggering an event, that trigger might never be reached.

Summary of Findings

The above analyses identify problematic cases in which both GR models yield unintended outcomes. The R&G model is relatively consistent and easy to understand but, faced with an apparently "irrational" agent, it oscillates between goals depending on the most recent observation. Used with higher β values, it is also beguilingly decisive, able to return, in the zigzagging example (Table 2), $P(g_3) = 0.98$ then half a dozen steps later $P(g_1) = 0.99$! Additionally, the more irrational the agent, the more confidently the distribution points towards the most probable goal (Theorem 3.3), apparent in the looping example (Table 1 at v), most obviously for $\beta = 0.1$, where probabilities start from a lower (flatter) base.

V&K, on the other hand, appears inconsistent and indecisive. Even when the agent seems clearly on the optimal path towards a particular goal (Table 2 at v), it assigns that goal a probability less than 0.5, scarcely outscoring its much more suboptimal competition. Moreover, faced with an irrational agent that seems not to be targeting any particular goal, it is biased to prefer the most distant/expensive goal, no matter where the agent is currently located. At first, it oscillates (note at w, it has "swung" from g_1 to prefer g_3, in agreement with R&G). As the path becomes more suboptimal, however, the ratio on which V&K depends becomes so diluted that the most distant/costly-to-reach goal (which supplies the largest numerator) again dominates (Table 2, locations x, y and z).

In fairness, both the above accounts were developed under the assumption of rationality. It is a "soft" assumption, however, in that we aim for a GR framework able to accommodate suboptimal behaviour. Indeed, both models derive their probability distributions based on the degree of suboptimality that they encounter. It appears, therefore, that rationality ought to be accommodated in the framework natively; and that is our objective in this paper.

3.3 Measuring the Degree of Irrationality

We have seen that probabilities generated by V&K's formula (3), when confronted by an even marginally suboptimal plan, can seem illogical in the way that it biases towards the most distant goal. Nonetheless, the *score* on which the probabilities are based degrades (behind the scenes) in an interesting and useful way.

The ratio $optc(s_0, g) \div optc(s_0, \vec{o}, g)$ used under the V&K model balances optimal cost from start to goal against optimal cost through the observations. Thus, a perfectly rational observed plan, where $optc(s_0, \vec{o}, g) = optc(s_0, g)$, yields a score of 1; but as the observed behaviour becomes increasingly erratic (i.e., suboptimal for *all* goals, that is, "strictly less rational"), the denominator *increases* (for all goals) while the numerator (for all goals) remains the same.

Note that if a plan is optimal (or close to optimal) for *some* goal, it is not an erratic or irrational one, and the maximum score approaches 1. Only when the plan is *suboptimal for all goals* is the maximum score diminished. It turns out, then, that the maximum score at any point in the plan provides a good measure of the degree to which optimality has (in general) become "diluted".

Definition 3.5. Relative to a GR problem, $\mathcal{P}$, the ***rationality measure (RM)*** is given by:

$$RM(s_0, G, \vec{o}) = \max_{g \in G} \frac{optc(s_0, g)}{optc(s_0, \vec{o}, g)}. \tag{4}$$

Notice that the RM is based on "strictly less rational" behaviour (Definition 3.1), a more general, cumulative measure than the notion of being "uniformly less rational" (Definition 3.2). Although it measures suboptimality across all goals, observations are not necessarily suboptimal for each goal by the same amount. By taking the maximum score, we always assess rationality based on the "best" possible interpretation. So, given two observation sequences $\vec{o}$ and $\vec{o}\,'$, where $\vec{o}\,'$ is uniformly less rational than $\vec{o}$, $RM(s_0, G, \vec{o}\,') < RM(s_0, G, \vec{o})$ (the more irrational the observations, the lower *all* the scores become; so the lower the *maximum* score becomes). Thus, the RM for *uniformly less rational* observations is always lower. The reverse does not apply, however. Observations with a lower RM are *not* always uniformly less rational. "Uniformly less rational" observations involve the same amount of unnecessary work for all goals; observations with a low RM also do a lot of unnecessary work but may still be tracking (albeit suboptimally) towards one particular goal.

Our model for GR (below) uses the RM "on-the-fly" to provide a snapshot of the agent's degree of rationality, based on their immediate history but the measure has other potential uses. For example, it could also be used from problem to problem, as follows. Once the RM for a particular agent has been established (on the basis of the current, or past, problem), it provides a means of predicting how suboptimal that agent's behaviour is likely to be in future. Though beyond the scope of this paper, this may have an impact on the sort of (suboptimal) planner that might be used to generate plans with which to compare observations, should the same agent be encountered a second time. If the RM is very low (i.e., highly suggestive of irrationality) then this might be used to flag the need for additional (perhaps human) surveillance on future sightings.

4 A SELF-MODULATING APPROACH TO GR

We now present our *self-modulating* account which uses the RM, in combination with a non-sigmoidal variation of R&G, to lift the rationality assumption.

Our objective is to obtain a probability distribution—a solution to a GR problem $\mathcal{P} = \langle \mathcal{D}, O, s_0, G, \vec{o}, Prob \rangle$—that preserves the intuition behind R&G that the lower the cost difference, the higher the probability but which modulates its level of confidence relative to the degree of rationality observed so far.

To achieve this, we propose the following formula:

$$P(G \mid \vec{o}) = \alpha \cdot \frac{1}{e^{\beta \; costdif(s_0, \vec{o}, g)}}, \tag{5}$$

where $\beta = RM(s_0, G, \vec{o})^{\gamma}$, and γ is a positive constant.

Formula (5) maintains an R&G-like awareness of the goal the agent seems to be approaching but, as the agent becomes irrational with respect to all goals (i.e., apparently cost-insensitive and possibly deceptive), the formula self-regulates and lowers its level of confidence accordingly. By doing so, it accommodates the rationality assumption in the process of performing GR. More concretely, while the agent behaves rationally (with respect to at least one of the goals), a confident prediction is returned, at the limit of those we have seen from R&G ($\beta = 1$); but the more irrational the agent becomes (i.e., their observed behaviour is increasingly suboptimal with respect to all goals in the domain), a less confident prediction is given, resembling the more subdued distributions of V&K or R&G ($\beta = 0.1$) (i.e., with lower β value).

Importantly, our formula (5) substitutes for R&G's Boltzmann equation a non-sigmoidal distribution (which does not suffer from the discrepancy captured by Theorem 3.3). It does, however, draw on a seldom-discussed feature of the R&G model: the β parameter.

4.1 The rate parameter in R&G

The solution to a GR problem under R&G is achieved using the Boltzmann probability distribution—formula (2)—tempered by a *rate parameter* β. As seen in Tables 1 and 2, while the value of β makes no difference to the relative *ranking* of goals within a probability distribution, it does have considerable impact on the *shape* of that distribution. Indeed, as briefly discussed in Ramirez's PhD thesis [11, p.63] (though mostly ignored in the papers):

> *This [parameter] allows plan recognition system developers to soften the implicit assumption of the agent being rational as in preferring those plans that minimize their total cost. The smaller the value of β the more will the distribution resemble a uniform distribution ...*

Thus, formula (2) already includes a parameter to control the level of confidence in the observed agent's rationality; but the choice of a value for β (given a value of 1 in the thesis and in code linked from [12]) is left to be set by the GR system developer, presumably on the basis of domain knowledge or special information about the particular agent under observation. Our approach, in Equation (5), is for the formula to self-adjust this parameter "on-the-fly" based on the RM, which, recall, we derive by maximising the score (not the probabilities) from V&K.

4.2 Properties of the Self-Modulating Formula

Formula (5), via the now dynamic β, synthesises the accounts discussed above to achieve the following properties and features.

(1) In place of the Boltzmann, this exponential distribution precisely enforces the intuition that the lower the cost difference, the higher the probability. Furthermore, unlike formula (2), it is closed under scaling and so returns consistent probability values when one sequence of observations is strictly less rational than another; also, it guarantees probabilities (before modification by the β parameter) always at the limit of those calculated under formula (2) (see proof of Theorem 3.3).

(2) The β parameter self-adjusts by reference to the agent's current degree of suboptimality, given by RM (formula 4).

(3) The confidence parameter γ regulates how quickly confidence should drop if irrational behaviour is detected. If γ is high, the less suboptimal the observations need to be before the probability distribution flattens out.

(4) Our formula (5) ostensibly requires three calls to a planner per goal—for $optc(s, \vec{o}, g)$, $optc^{\neg}(s, \vec{o}, g)$ and $optc(s, g)$—whereas the R&G distribution requires only the first two. However, the additional call (required for the RM formula 4) depends on the domain, not on the observations, so can be precalculated and cached. Taking this approach, self-modulation can be achieved without time penalty. (Our main focus was nonetheless on capturing the intended meaning more accurately, above achieving computational efficiency.)

In the following, $P(\cdot)$ stands for our self-modulating formula; $\beta_{\vec{o}}$ represents the β value for observations $\vec{o}$ from Equation (5); and $P_{RG}^{\beta=x}$ represents Equation (2) with $\beta = x$.

First, we formalise the observation at (1) above, which follows from $\lim_{a\to\infty}[1/(1+a) \div 1/a] = 1$ in the proof of Theorem 3.3.

Observation 1. $\lim_{costdif(s_0,\vec{o},g)\to\infty} P_{RG}^{\beta=\beta_{\vec{o}}}(\cdot) \div P(\cdot) = 1.$

Next, the more rational an agent's behaviour (i.e., the observation sequence $\vec{o}$), the higher the β in our account and, thus, the more closely probabilities approach those at the limit of R&G ($\beta = 1$).

Theorem 4.1. *Let $\vec{o}, \vec{o}\,' \in O^*$ be two observation sequences such that $\vec{o}\,'$ is strictly less rational than $\vec{o}$. Then,* $1 \geq \beta_{\vec{o}} > \beta_{\vec{o}\,'}$.

Proof. (Sketch) β is based on the RM (formula 4) which maximises the ratio at formula (3). Across goals, numerators remain constant. As observed costs increase, all denominators increase and all values decrease: therefore the maximum must decrease. □

Now, when an agent behaves fully rationally—that is, navigates optimally with respect to some goal—our probability distribution is at the limit of R&G ($\beta = 1$) because, when observations conform to optimal behaviour, $\beta = 1$ in Equation (5). Moreover, see next that even when our account diverges from R&G ($\beta = 1$), it still maintains the same relative rankings across goals.

Theorem 4.2. *For all observations $\vec{o} \in O^*$ and goals $g_1, g_2 \in G$, $P(g_1 \mid \vec{o}) > P(g_2 \mid \vec{o})$ iff $P_{RG}(g_1 \mid \vec{o}) > P_{RG}(g_2 \mid \vec{o})$.*

Proof. (Sketch) The differences between formulas (2) and (5) have no impact on probability rankings. Specifically:

(1) subtracting +1 from the denominator of *every* score does not change their relative order; and

(2) β is a multiplicative constant. Changing its value (including by the introduction of γ) effects a monotonic transformation, again maintaining the relative order of probabilities. □

Thus, our formula is aligned with the underlying assumption of the R&G framework with respect to the rationality of the observed agent and never alters the qualitative outcome: *goal rankings are maintained.* Critically, though, as the following important result states, the more erratic the observations (i.e., the more suboptimal for *all* goals), the more even the probability distribution becomes.

Theorem 4.3. *Let $\vec{o}, \vec{o}\,' \in O^*$ be two observation sequences such that $\vec{o}\,'$ is strictly less rational than $\vec{o} \in O^*$. Then, for every two goals $g_1, g_2 \in G$ such that $P(g_1 \mid \vec{o}\,') \neq P(g_2 \mid \vec{o}\,')$ (i.e., whenever the two goals are distinguishable):*

$$|P(g_1 \mid \vec{o}\,') - P(g_2 \mid \vec{o}\,')| < |P(g_1 \mid \vec{o}) - P(g_2 \mid \vec{o})|.$$

Table 3: Probabilities Revisited.

	R&G ($\beta = 0.1$)			Self-Mod			$\gamma = 2$
	g_1	g_2	g_3	g_1	g_2	g_3	β
s_2	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.2642
s_3	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.1196
v_1	***0.4200***	0.3342	0.2458	**0.9842**	0.0156	0.0001	1
v_2	***0.4656***	0.3237	0.2106	**0.5752**	0.2906	0.1342	0.1649
v_3	***0.4789***	0.3196	0.2015	**0.4295**	0.3283	0.2422	0.0649
v_{10}	***0.4820***	0.3186	0.1994	**0.3383**	0.3335	0.3281	0.0050

(a) Loopy Paths.

	R&G ($\beta = 1$)			Self-Mod			$\gamma = 2$
	g_1	g_2	g_3	g_1	g_2	g_3	β
v	**0.9694**	0.0303	0.0003	**0.9842**	0.0156	0.0001	1
w	0.0008	0.0156	**0.9835**	0.1267	0.2458	**0.6275**	0.2262
x	0.3368	**0.6050**	0.0581	0.3514	**0.3886**	0.2500	0.1716
y	**0.9951**	0.0049	4.5e-05	**0.6018**	0.2674	0.1308	0.1526
z	0.0200	0.3734	**0.6066**	0.2894	0.3513	**0.3637**	0.0715

(b) Zigzagging Path.

Proof. (Sketch)

(1) Since $\vec{o}\,'$ is strictly less rational than $\vec{o}$, by Definition 3.1, for all $g_1 \in G$, $optc(s_0, \vec{o}\,', g_1) > optc(s_0, \vec{o}, g_1)$.
(2) From Theorem 4.1, β in Formula (5) is reduced when the cost of the observation sequence increases.
(3) When β is reduced, a monotonic transformation diminishes the difference between probabilities. □

4.3 Comparison of Results

In Tables 3a and 3b, we compare the probabilities returned by our self-modulating formula (5) with the original "static" R&G formula (2). Referring to Table 3b, observe that, at location v, when the agent appears to be on an optimal path to g_1 (though on a suboptimal path to g_2 and g_3), "Self-Mod" (5) maintains $\beta = 1$ and therefore yields a confident prediction. But, as the path becomes increasingly suboptimal, the distribution obtained by formula (5) evens out so that, by location z, the most and least likely goals are separated by just 0.07 (compared with 0.58 using static R&G).

Notice that Self-Mod always maintains the same rankings as R&G but, referring now to Table 3a, we clearly see that whereas a looping path causes Self-Mod to return a less confident prediction (the more the path loops, the more the distribution flattens out), the opposite is true of R&G, which counter-intuitively increases in confidence with every loop. Thus, using formula (5), the validity of the assumption with respect to the rationality of the observed agent has been accounted for.

Thus, our self-modulating formula provides the performance we set out to achieve: as the agent becomes more erratic (suboptimal), it yields a distribution closer to uniform. Practically, faced with apparently irrational behaviour, a GR system using our approach judges goals more equally, displaying a reduced level of confidence.

5 RELATED WORK

We briefly reflect on other extensions to, and adaptations of, [12]. First, as previously noted, the term $optc^{\neg}(s_0, \vec{o}, g)$ in cost difference formula (1) can be replaced by $optc(s_0, g)$. We have shown elsewhere that this alternative formulation, also considered by E-Martin et al. [3], yields an identical result to the original term in all cases barring those where observations conform to the optimal way of attaining *multiple* goals and the *only* optimal way of attaining one of them [9]. The simpler formula is less demanding computationally (no need to reason negatively about observations and the term is reusable if probabilities are to be checked multiple times) and more robust (if the initial state s_0 is the only observation, it yields prior *Prob*).

In relation to apparent irrationality, like us, Sohrabi et al. [14] deal with unreliable observations. Their focus, however, is not suboptimality but the possibility that observations may be either noisy or missing. Interestingly, their solution—when applied to increasingly suboptimal paths—skews results so as to *penalise* the most distant goal (which V&K favours). This is because an irrational path is "noisy" for all goals to approximately the same extent; but the optimal path to a distant goal inevitably has more missing observations than the optimal path to a goal nearby.

In recent work, Vered and Kaminka [16] introduce heuristics to speed up the recognition process. As usual, rationality is assumed: if an agent behaves irrationally with respect to a particular goal (e.g., in a path-planning context, by turning away from it), that goal is pruned from future consideration. In [17], the authors explicitly consider rationality/irrationality and observe that humans performing goal recognition are particularly susceptible to its dictates (more so, in fact, than their formula). Indeed, as Jian et al. [5] have shown, when drawing a path with the intent to deceive, people immediately subvert the rationality assumption, typically settling on massively suboptimal (e.g., spiralling or zigzagging) paths.

6 CONCLUSION

We have lifted perhaps the strongest assumption in current state-of-the-art approaches to GR, that of the rationality of the observed agent. As a result, we can handle agents ranging from the strictly rational to the arbitrarily irrational in a principled manner.

We first analysed two well-respected contemporary approaches and identified situations in which they yield unexpected results (when faced with plans that appear irrational). By synthesising both approaches via a measurement of the agent's expected degree of suboptimality, we devised an alternative model for GR. Importantly, the synthesis is principled and conceptually simple. The proposed model respects the intuition that, typically, a plan's probability is inversely proportional to its cost, but degrades gracefully if the underlying assumption of rationality is compromised.

In future work, there is scope to consider how confidence could be restored if, after a period of irrationality or erratic behaviour, the observed agent seems once again to be "back on track".

ACKNOWLEDGEMENTS

We thank our reviewers for their suggestions and acknowledge support from the Australian Government Research Training Program Scholarship.

REFERENCES

[1] Elisheva Bonchek-Dokow and Gal A. Kaminka. 2014. Towards computational models of intention detection and intention prediction. *Cognitive Systems Research* 28 (2014), 44–79.

[2] Sandra Carberry. 2001. Techniques for plan recognition. *User Modeling and User-Adapted Interaction* 11, 1-2 (2001), 31–48.

[3] Yolanda E-Martin, Maria D R-Moreno, David E Smith, et al. 2015. A fast goal recognition technique based on interaction estimates. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI).* 761–768.

[4] Jaakko Hintikka. 1962. Knowledge and Belief: An Introduction to the Logic of the Two Notions.

[5] Jiun-Yin Jian, Toshihiko Matsuka, and Jeffrey V. Nickerson. 2006. Recognizing deception in trajectories. In *Proc. of the Cognitive Science Society.* 1563–1568.

[6] Peta Masters and Sebastian Sardina. 2017. Cost-based goal recognition for path-planning. In *Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS).* 750–758.

[7] Peta Masters and Sebastian Sardina. 2017. Deceptive Path-Planning. *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)* (2017), 4368–4375.

[8] Peta Masters and Sebastian Sardina. 2018. Cost-Based Goal Recognition for the Path-Planning Domain.. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI).* 5329–5333.

[9] Peta Masters and Sebastian Sardina. 2019. Cost-Based Goal Recognition in Navigational Domains. *Journal of Artificial Intelligence Research* 64 (2019), 197–242.

[10] Ramon Fraga Pereira, Nir Oren, and Felipe Meneguzzi. 2017. Landmark-based heuristics for goal recognition. In *Proceedings of the National Conference on Artificial Intelligence (AAAI).* 3622–3628.

[11] Miquel Ramirez. 2012. *Plan Recognition as Planning.* Ph.D. Dissertation. Univeristat Pompeu Fabra.

[12] Miquel Ramirez and Hector Geffner. 2010. Probabilistic plan recognition using off-the-shelf classical planners. In *Proceedings of the National Conference on Artificial Intelligence (AAAI).* 1121–1126.

[13] Maayan Shvo, Shirin Sohrabi, and Sheila A McIlraith. 2018. An AI Planning-Based Approach to the Multi-Agent Plan Recognition Problem. In *Advances in Artificial Intelligence: 31st Canadian Conference on Artificial Intelligence, Canadian AI 2018, Toronto, ON, Canada, May 8–11, 2018, Proceedings 31.* 253–258.

[14] Shirin Sohrabi, Anton V. Riabov, and Octavian Udrea. 2016. Plan Recognition as Planning Revisited.. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI).* 3258–3264.

[15] Gita Sukthankar and Katia Sycara. 2005. A cost minimization approach to human behavior recognition. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI).* 1067–1074.

[16] Mor Vered and Gal A. Kaminka. 2017. Heuristic Online Goal Recognition in Continuous Domains, In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI). *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 4447–4454.

[17] Mor Vered, Gal A. Kaminka, and Sivan Biham. 2016. Online goal recognition through mirroring: Humans and agents. In *Conference on Advances in Cognitive Systems.*