How Tacit Knowledge Guides Action

Ya'akov Gal
MIT CSAIL, and
Harvard University SEAS

Rajesh Kasturirangan
National Institute of
Advanced Studies
Bangalore, India, and
MIT CSAIL

Avi Pfeffer
Harvard University SEAS

Whitman Richards
MIT CSAIL

Abstract

Natural Intelligence is based not only on conscious procedural and declarative knowledge, but also on knowledge that is inferred from observing the actions of others. This knowledge is tacit, in that the process of its acquisition remains unspecified. However, tacit knowledge (and beliefs) is an accepted guide of behavior, especially in unfamiliar contexts. In situations where knowledge is lacking, animals act on these beliefs without explicitly reasoning about the world or fully considering the consequences of their actions. This paper provides a computational model of behavior in which tacit beliefs play a crucial role. We model how knowledge arises from observing different types of agents, each of whom reacts differently to the behaviors of others in an unfamiliar context. Agents' interaction in this context is described using directed graphs. We show how a set of observations guide agents' knowledge and behavior given different states of the world.

Introduction

How do we infer the proper behavior in unfamiliar or sensitive contexts? Suppose you are in a line, waiting to greet a foreign dignitary. Do you bow, or do you shake hands? What is the appropriate protocol? Similarly, at a formal dinner, what would be the proper etiquette – especially in an unfamiliar culture? Even the act of extended eye contact may be a faux pas if you are facing a hostile gang or a malevolent dictator. Studies show that individuals (and presumably lower animals as well) act appropriately in the world without an explicit representation of all of its features (Brooks 1991; Gibson 1977; Bonabeau, Dorigo, & Theraulaz 1999).

Tacit beliefs — in combination with information from other sensory modalities— are often invoked in new contexts. For example, one general class of tacit beliefs revolves around judgments of typicality and mimicry, i.e., "copy someone who looks like he knows what he is doing". When waiting in line to meet the foreign dignitary, one possible strategy is to follow the actions of someone whose dress and mannerism convey familiarity about the proper etiquette. Here typicality is assessed by looking at dress and mannerisms which is information provided by perception. Mimicry and typicality judgments are tacit knowledge since they are not stated formally. We are guided mostly by what other people do, not by rigorous analysis and reasoning (Minsky 1998).

Tacit beliefs allow agents to acquire knowledge and infer accepted modes of behavior in new contexts without fully reasoning about all the possible consequences of their actions. Suppose that a driver that is unfamiliar with the traffic laws in a particular country is waiting to turn right behind a line of cars and can only see a school bus in front. If the school bus turns right on a red light, the driver of the car behind it may choose to follow its actions and do the same. However, this driver may choose not to follow a beat-up sedan. This is because this driver assumes the school bus to be following the local rules, while the driver of the beat-up sedan is believed to be reckless. This paper presents preliminary work towards a computational theory (Marr 1982; Agre & Rosenchein 1996) of the way tacit beliefs affect the way agents interact with each other in a network.

A Basic Model

Initially we will consider binary rules and actions. A rule R_+ states that an action is legal (e.g., turning right on red). Rules are assumed to be consistent in the world, so R_+ and R_- cannot both hold at the same time.

A knowledge of a rule R_+ by agent i is denoted as $K_i(R_+)$. For each rule one of the following must hold for each agent i: i knows the action is legal ($K_i(R_+)$); i knows the action is illegal ($K_i(R_-)$); or i does not know whether the action is legal or illegal.

Knowledge in our model is assumed to be *correct*, such that $K_i(R_+) \longrightarrow R_+$. It also follows that knowledge is also consistent, such that $K_i(R_+) \longrightarrow \overline{K_i(R_-)}$. (And similarly for R_-). Note that the converse does not hold. For example, $\overline{K_i(R_-)}$ does not imply that R_- hold. Intuitively, not knowing whether it is illegal to turn right on red does not imply that this action is illegal.

We use a + symbol to mean that an action was carried out and a − symbol to mean that an action was not carried out. We use the notion of *types* to distinguish between what different agents do, given their knowledge about rules in the world. A type essentially refers to an agent's strategy, and this strategy is specific to each type of agent. [1] I For example, consider the action of turning right on a red light. We can define the following types for agent i:

- t_1 (conservative). Choose action − (Never turn right on red)
- t_2 (law abiding). Choose action + if $K_i(R_+)$ (Turn right on red only if you know it is legal)
- t_3 (risk taker). Choose action + if $\overline{K_i(R_-)}$. (Turn right as long as you don't know it is illegal)
- t_4 (reckless). Choose action + (Always turn right on red)

The way in which agents interact with each other in our model is defined by a directed network called an *interaction graph*. Each node in the network represents a type of agent. We assume that agents have tacit knowledge of the types of agents they interact with, and they use this knowledge as a surrogate for what is true in the world. In an interaction graph, an edge from agent t_i to agent t_j means (1) that agent t_j knows the type of agent t_i, and (2) that agent t_j can observe the action of type t_i. In the traffic example, the following network represents a possible instantiation of an interaction graph in which there is a line of cars, where an agent of type t_3 is waiting behind an agent of type t_1 who is waiting behind an agent of type t_2, etc...

$$t_4 \rightarrow t_2 \rightarrow t_1 \rightarrow t_3$$

The following table summarizes what actions are taken by each agent type, given its neighbors. We list the actions for a row agent i given that its neighbor is the column agent j that chooses action + (left entry in a column) and action − (parenthetical right entry in a column). A $\emptyset$ symbol denotes a counter-factual event — an action that cannot be chosen by a given type under the circumstances.

	t_1	t_2	t_3	t_4
t_1	$\emptyset$ (−)	− (−)	− (−)	− ($\emptyset$)
t_2	$\emptyset$ (−)	+ (−)	− (−)	− ($\emptyset$)
t_3	$\emptyset$ (+)	+ (+)	+ (−)	+ ($\emptyset$)
t_4	$\emptyset$ (+)	+ (+)	+ (+)	+ ($\emptyset$)

We say that agent i *conveys knowledge* if its actions provide information about rules in the world, or about the knowledge of other agents about rules in the world. Consider for example an edge (t_4, t_2) in a possible interaction graph for the traffic example, meaning that an agent of type t_2 is waiting behind an agent of type t_4. Suppose that the t_2 agent does not know whether R_+ or R_- hold. The t_4 agent always turns right, so it doesn't convey additional knowledge to the t_2 agent. Since t_2 is law-abiding, it will choose action −. In contrast, consider an edge (t_4, t_3) and that $\overline{K_i(R_-)}$ holds for the agent of type t_3. In this case, the t_3 agent will choose action +, because the actions of the t_4 have not revealed that R_- holds. To summarize, we can infer the following about types $t_1, \ldots, t_4$ in an interaction graph.

- Types t_1 and t_4 never convey knowledge because the strategies of these types do not depend on knowledge of a constraint.
- Type t_3 conveys $K_i(R_-)$ when it is observed to do −, and conveys $\overline{K_i(R_-)}$ when it is observed to do +.
- Type t_2 conveys $K_i(R_+)$ when it is observed to do + and conveys $\overline{K_i(R_+)}$ when it is observed to do −.

We can now state the following:

Theorem 1. *Let C be an interaction graph. An agent j of type t_2 in C will choose action + if and only if the following hold:*

- *There is an agent i in C such that $K_i(R_+)$ holds.*
- *There is a directed path in C from i to j that passes solely through agents of type t_2.*

Similarly, an agent t_3 will do − when $K_i(R_-)$ holds and there is a path from i to j that passes solely through agents of type t_3.

This theorem is easy to prove. Take for example a path from R_+ to an agent of type t_2. Any agent along this path that is not of type t_2 will not convey knowledge of R_+ to t_2, and therefore t_2 will do −, as specified by the table. The argument is similar for a path from R_- to an agent of type t_3.

A corollary of Theorem 1 is that any knowledge that is conveyed to an agent in an interaction graph is consistent. To see this, suppose both $K_i(R_+)$ and $K_i(R_-)$ hold for some agent i in an interaction graph. By Theorem 1, both R_+ and R_- will be descendants of i, which cannot be the case, because rules are consistent.

Using the theorem, we can induce a mapping from any agent i and interaction graph C to a knowledge condition for i, stating that i knows the rule is legal ($K_i(R_+)$), i knows the rule is illegal ($K_i(R_-)$), or that i does not know ($\emptyset$). Note that agents convey knowledge based on their types and are not perfect reasoners given the structure of the graph. For example, consider a network $t_2 \rightarrow t_3$, in which $K_i(t_2)$ holds. Agent t_2 would do −, and according to the table, agent t_3 would do +, because it cannot infer whether $K_i(R_-)$ holds for the agent in front of it. However, if t_3 knew the structure of the graph it would be able to infer $K_i(R_-)$ from the action of t_2 and the fact that t_2 can observe R_-.

A Generative Model of Types

A type is a mapping from a set of possible knowledge predicates for an agent to an action. The number of

[1] Our use of types in this work is distinguished from the traditional usage of this term as representing agents' private information in game theory.

possible knowledge predicates depend on the number of rules, and the number of values each rule can take. In the traffic example, either $K_i(R_+)$ or $\overline{K_i(R_-)}$ can hold for each agent (and similarly for R_-). We consider any instantiation of these conditions that is consistent in the world as a valid knowledge predicate. The possible knowledge predicates the traffic example are as follows:

1. $\{K_i(R_+), \overline{K_i(R_-)}\}$
2. $\{K_i(R_-), \overline{K_i(R_+)}\}$
3. $\{\overline{K_i(R_+)}, \overline{K_i(R_-)}\}$

Note that the set $\{K_i(R_+), K_i(R_-)\}$ cannot occur because of the consistency condition of our model.

Thus, an agent of type t_3 (risk taker) chooses action + if (1) and (3) hold, while an agent of type t_2 (law abiding) chooses action + solely if (1) holds. Types t_1 and t_4 choose action + and action − respectively for all possible sets of knowledge predicates.

In general, if there are n binary rules, there are three possible sets of knowledge predicates for each rule, and the total number of possible sets is thus 3^n. A type is a mapping from each of these sets to an action + or −, so the number of possible types is $2^{(3^n)}$, which is doubly exponential in the number of rules. However, using Theorem 1, we can enumerate determine the knowledge and the actions of particular agents without having to enumerate all types.

We now show that $t_1, \ldots t_4$ is not an arbitrary subset of the $2^3 = 8$ possible types in the traffic example. Let S be a complete order over knowledge predicates. We say that a knowledge predicate i *dominates* j if $i \succ j$ in S. In the traffic example, we impose an ordering (2) $\succ$ (3) $\succ$ (1). Intuitively, this ordering represents a degree of severity. Knowing an action is illegal is the highest degree of severity, whereas knowing that an action is legal is the lowest degree.

We then say that a type t_i is *monotonic* if for any two knowledge predicates K_1, K_2 such that $K_1 \succ K_2$ the following holds: if t_i chooses action a_1 in K_1, and $a_1 \in K_1, a_2 \in K_2$ then $a_1 \succ a2$.

It turns out that all of the types in our traffic example are monotonic with the ordering imposed above. For example, it is sufficient to characterize a risk taker type t_3 as doing − under knowledge condition (2), in which R_- is known and choosing to do + under knowledge condition (3), in which R_- is not known. Given the ordering we've imposed above, it follows that t_3 will choose to do "+" under knowledge condition (1), in which R_+ is known. As an example of a non-monotonic type, consider a "malicious" agent that chooses action + solely under knowledge condition (2). This agent chooses + when it knows R_-, and chooses − when it knows R_+. Another example is one that drives at least 45 MPH even if it does not know the legal speed limit, never drives above 65 MPH, and drives at the maximal speed x it knows to be legal where $45 < x < 65$.

Multiple Actions

We now turn to the case in which there are multiple possible actions to consider. Consider the speed of the car as a discrete variable, where each value represents that the speed of the car is within a possible interval of speeds, say 10 MPH. Suppose that the minimal and maximal speed of a car are min and max, respectively. A rule in this example, denoted R^y denotes that the maximal driving speed is y. There is a natural ordering over knowledge predicates in this scenario, mainly that $K_i(R^y) \succ K_i(R^z)$ for every $z \leq y$. We need only consider types that are monotonic with respect to this ordering. Thus, a type t_i^y implies that agent i drives at speed y or below. The set of monotonic types we introduced for the binary case generalize to the multi-value case. A law abiding agent type drives at speed y or below if it knows R^y. Otherwise it drives at a minimal speed of min. A risk taker agent type drives at speed z as long as it does not know R^y. If it knows R^y, it drives at speed z if $z \leq y$ and at speed y if $z > y$.

The following table summarizes what actions are taken by each agent type, given its neighbors. We list the actions for a row agent i given that its neighbor is the column agent j that is observed to be driving at speed y. We use notation "RT" to refer to a risk taking agent and "LA" to refer to a law abiding agent.

	LA	RT
LA	$\leq y$	min
RT	$\geq y, \leq z$	$\leq y, \leq z$

A law abiding agent always conveys that the speed limit is y. Thus a risk taker that is observing a law abiding agent in the network that is driving at speed y will always drive at a speed that is greater or equal to y. However, a risk taker never conveys the speed limit, and a law abiding agent that observes it can only drive at the minimal speed.

We can also extend Theorem 1 to fit this case. A law abiding agent type t will drive at speed y (or below) if there is an agent i in the interaction graph for which $K_i(R^y)$ holds, and a path from this agent to t that is composed solely of law abiding agents. (and similarly for a risk taking agent).

Discussion

The interaction graph in our traffic example above was a simple directed chain, with four different types of nodes arranged in an arbitrary order. There are many other forms of interaction graphs. Obvious cases are a leader of a group (connected graph), or when everyone can see everyone else (complete graph), a bipartite graph, random graphs, etc. With only 8 nodes, there are 10^{12} possible forms (Harary 1969). Note that Theorem 1, however, is general and lays out a condition where knowledge of the graph form and node types can lead to a correct action. Theorem 1 illustrates the larger claim

that whenever tacit knowledge is governed by underlying principles such as the global consistency of knowledge, agents have enough information to do the right thing. In our case, the rules and types were all categorical. Other cases, however, may require probabilistic inferences. We will discuss some of these extensions and conditions for correct inference.

The other interesting computational element in our model is the use of monotonicity as a constraint to address a difficult inverse problem, namely, to infer the right action based on observations of other agents behavior. In general, there cannot be a unique solution to the problem of right action. However, the assumption that people fall into types constrains the inference procedure. However, types alone are not enough. We need to assume that both types of agents and the set of observations are structured in a partial ordering. In our traffic example, types are ordered according to their response to punishment (reckless to conservative). We also assume that agents knowledge is ordered as well, i.e., that no one ever knows less than she knew before so observations always increase knowledge. These two are distinct partial orders but they are related, for whenever there is a scale of punishment, it helps if observations help rather then hurt.

Finally, although the inference process may in some cases be complex, our main intent is to present a simple representation for evaluating how tacit beliefs and knowledge can lead to correct actions in unfamiliar contexts. The fact that tacit knowledge is largely implicit, suggests how lower animals as well as humans can learn appropriate behaviors without directed examples and tutors.

References

Agre, P., and Rosenchein, S. 1996. *Computational Theories of Interaction and Agency*. MIT Press.

Bonabeau, E.; Dorigo, M.; and Theraulaz, G. 1999. *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press.

Brooks, R. A. 1991. Intelligence without representation. artificial intelligence. *Artificial Intelligence* 47:139–159.

Gibson, J. J. 1977. The theory of affordance. In Shaw, R., and Bransford, J., eds., *Perceiving, Acting, and Knowing*.

Harary, F. 1969. *Graph Theory*. Addison-Wesley.

Marr, D. 1982. *Vision*. Freeman and Co.

Minsky, M. 1998. *The Society of Mind*. Simon and Schuster. 302–303.