Cognitive Science 30 (2006) 147–180

Formalization and Analysis of Reasoning by Assumption

Tibor Bosse[a], Catholijn M. Jonker[b], Jan Treur[a]

[a]*Department of Artificial Intelligence, Vrije Universiteit Amsterdam, The Netherlands*
[b]*Nijmegen Institute for Cognition and Information, Radboud Universiteit Nijmegen, The Netherlands*

Received 6 October 2004; received in revised form 28 April 2005; accepted 28 June 2005

Abstract

This article introduces a novel approach for the analysis of the dynamics of reasoning processes and explores its applicability for the reasoning pattern called *reasoning by assumption.* More specifically, for a case study in the domain of a Master Mind game, it is shown how empirical human reasoning traces can be formalized and automatically analyzed against dynamic properties they fulfill. To this end, for the pattern of reasoning by assumption a variety of dynamic properties have been specified, some of which are considered characteristic for the reasoning pattern, whereas some other properties can be used to discriminate among different approaches to the reasoning. These properties have been automatically checked for the traces acquired in experiments undertaken. The approach turned out to be beneficial from two perspectives. First, checking characteristic properties contributes to the empirical validation of a theory on reasoning by assumption. Second, checking discriminating properties allows the analyst to identify different classes of human reasoners.

Keywords: Human reasoning; Assumptions; Formalization; Empirical traces; Formal analysis; Dynamics; Master Mind

1. Introduction

Practical reasoning processes are often not limited to single reasoning steps, but extend to traces or trajectories of a number of interrelated reasoning steps over time. In the analysis of such reasoning processes, dynamic aspects play an important role. Examples of such dynamic aspects are posing reasoning goals, making assumptions, and evaluating assumptions. As a consequence, such reasoning processes cannot be understood, justified, or explained to others without taking into account these dynamic aspects. Therefore, the main goal of this article is to present a novel approach for the analysis of the dynamics of reasoning processes. This ap-

Correspondence should be addressed to Tibor Bosse, Department of Artificial Intelligence, Vrije Universiteit Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands. E-mail: tbosse@cs.vu.nl

proach is based on a combination of formal methods and human experiments. More specifically, it consists of a number of steps:

- First, a collection of *empirical data* is acquired, using an experiment in human reasoning.
- Next, the obtained transcripts are *formalized* using the Temporal Trace Language (TTL). This language was already shown to be a useful analysis tool for reasoning processes (Jonker & Treur, 2002).
- Next, a number of *dynamic properties* of reasoning processes are formalized using TTL. These can be divided into two categories: *characterizing properties* are expected to hold for all reasoning processes (e.g., "the process terminates"), whereas *discriminating properties* are expected to hold for some reasoning processes (e.g., "this particular reasoner uses the 'stepwise' strategy").
- After that, using an *automated checking* tool, it is investigated which dynamic properties hold for which transcripts. Such an analysis can be useful in two different ways. On the one hand, checking characterizing properties contributes to the validation of a theory on reasoning. On the other hand, checking discriminating properties helps to distinguish several types of transcripts from each other, thereby obtaining a classification of different reasoning strategies.
- Finally, *logical relations* are established among different dynamic properties, indicating how a number of properties together entail another (global) property. As is explained in Section 7, such logical relations play an important role in the analysis of empirical reasoning processes.

A more detailed description of the different steps of the approach is given in the remainder of this article.

As a second contribution, this article demonstrates how the analysis approach can be applied for a specific reasoning pattern in human problem solving called *reasoning by assumption.* This practical reasoning pattern involves a number of interrelated reasoning steps and uses in its reasoning states not only content information but also meta-information about the status of content information and about control. For this reasoning pattern, human reasoning protocols have been acquired, analyzed, formalized, checked on dynamic properties, and compared.

To obtain a specific case study in reasoning by assumption, the game of Master Mind was selected. This is a two-player game of logic, which was invented in 1970–1971 by Mordecai Meirowitz (Nelson, 2000). The goal of the game is to discover a secret code of three colored pegs, which can be obtained by making guesses and receiving information about the correctness of the guesses. Because of its protocol, the pattern of reasoning by assumption occurs frequently within this game. Therefore, the game of Master Mind (in a simplified version) will be the main case study within this article.

In Section 2 the underlying dynamic perspective on reasoning is discussed in more detail. Based on this perspective, a specific model for the pattern reasoning by assumption is presented, adopted from Jonker and Treur (in press). In Section 3, the temporal language TTL, used to express properties of reasoning processes, is introduced in detail. Next, in Section 4 it is shown how think-aloud protocols involving reasoning by assumption in the game of Master Mind can be formalized to reasoning traces. A number of the dynamic properties that have been identified for patterns of reasoning by assumption are shown in Section 5. For the acquired reasoning traces, the identified dynamic properties have been (automatically) checked. The results of these checks are provided in Section 6. In Section 7, it is shown how logical relations among dynamic properties at different abstraction levels can play a role in the analysis of empirical reasoning processes. Section 8 discusses the difference between human strategies and optimal strategies, and Section 9 is a conclusion. Appendix A contains the complete list of relevant dynamic properties that have been identified for the pattern of reasoning by assumption. Appendix B contains a number of additional logical relations among dynamic properties at different abstraction levels. Appendix C contains two example human transcripts and their formalization.

2. The dynamics of reasoning

In history, formalization of the cognitive capability to perform reasoning has been addressed from different areas and angles: philosophy, logic, cognitive science, artificial intelligence. Within philosophy and logic much emphasis has been put on the results (conclusions) of a reasoning process, abstracting from the process by which such a result is found: When is a statement a valid conclusion, given a certain set of premises? Within artificial intelligence, much emphasis has been put on effective inference procedures to automate reasoning processes. The dynamics of such inference procedures usually are described in a procedural, algorithmic manner; dynamics are not described and analyzed in a conceptual, declarative manner. Within cognitive science, reasoning is often addressed from within one of the two dominant streams: the syntactic approach (based on inference rules applied to syntactic expressions, as common in the logic-based approach (e.g., Braine & O'Brien, 1998; Rips, 1994), or the semantic approach (based on construction of mental models; e.g., Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Schroyens, Schaeken, & d'Ydewalle, 2001; Yang & Bringsjord, 2001; Yang & Johnson-Laird, 1999). Formalization and formal analysis of the dynamics within any of these approaches has not been developed in-depth yet. Recently, it was shown by Stenning and van Lambalgen (in press) that the syntactic and the semantic approach are not as mutually exclusive as they are often presented. They claim that by concentrating on the difference between inference rules and mental models, the field misses the important distinction between classical and defeasible logics. Their argument is that the real issue is often whether a person reasons *toward* an interpretation or *from* an interpretation to a conclusion.[1] In line with their claims, this article does not make any commitments to one of both approaches either. Instead, it introduces a generic methodology to analyze reasoning processes, no matter whether these are represented by rules or by models.

To understand a specific reasoning process, especially for practical reasoning in humans, the dynamics are important. In particular, for reasoning processes in natural contexts, dynamic aspects play an important role and have to be taken into account, such as dynamically posing goals for the reasoning, or making additional assumptions during the reasoning, thus using a dynamic set of premises within the reasoning process. Decisions made during the process—for example, on which reasoning goal to pursue, or which assumptions to make—are an inherent part of such a reasoning process. Such reasoning processes or their outcomes cannot be understood without taking into account these dynamic aspects.

The approach to the semantical formalization of the dynamics of reasoning presented in this section is based on the concepts' reasoning state, transitions among reasoning states, and reasoning traces: traces of reasoning states. Based on these concepts, in Section 2.4, a specific model for the pattern reasoning by assumption is presented, adopted from Jonker and Treur, (2003).

2.1. Reasoning state

A reasoning state formalizes an intermediate state of a reasoning process. It may include information on different aspects of the reasoning process, such as content information or control information. Within a syntactical inference approach, a reasoning state includes the set of statements derived (or truth values of these statements) at a certain point in time. Within a semantical approach based on mental models, a reasoning state may include a particular mental model constructed at some point in time, or a set of mental models representing the considered possibilities. However, additional meta-information can also be included in a reasoning state, such as control information indicating what is the focus or goal of the reasoning, or information on which statements have been assumed during the reasoning. Moreover, to be able to cover interaction between reasoning and the external world, part of the state of the external world is also included in a reasoning state. This can be used, for example, to model the presentation of a reasoning puzzle to a participant, or to model the participant's observations in the world. The set of all reasoning states is denoted by RS.

2.2. Transition of reasoning states

A transition of reasoning states, that is, an element $< S, S' >$ of $RS \times RS$, defines a step from one reasoning state to another reasoning state; this formalizes one reasoning step. A *reasoning transition relation* is a set of these transitions, or a relation on $RS \times RS$. Such a relation can be used to specify the allowed transitions within a specific type of reasoning. Within a syntactical approach, inference rules such as *modus ponens* typically define transitions between reasoning states. For example, if two statements

$$p, p \rightarrow q$$

are included in a reasoning state, then by a *modus ponens* transition, a reasoning state can be created where, in addition

$$q$$

is also included. Within a semantical approach, a construction step of a mental model, after a previous mental model, defines a transition between reasoning states. For example, if knowledge "if p then q" is available, represented in a mental state

$$[p], q$$

and, in addition, not-q is presented, then a transition may occur to a reasoning state consisting of a set of mental models

p, *q*; ~*p*, ~*q*; ~*p*, *q*

that represents the set of possibilities considered; a next transition may involve the selection of the possibility that fits not-q, leading to the reasoning state

~*p*, ~*q*

2.3. Reasoning trace

Reasoning dynamics or reasoning behavior is the result of successive transitions from one reasoning state to another. By applying transitions in succession, a time-indexed sequence of reasoning states $(\gamma_t)t \in T$ is constructed, where T is the time frame used (e.g., the natural numbers). A reasoning trace, created in this way, is a sequence of reasoning states over time, that is, an element of RS^T. *Traces* are sequences of reasoning states such that each pair of successive reasoning states in this trace forms an allowed transition, as has been defined under transitions. A trace formalizes one specific line of reasoning. A *set of reasoning traces* is a declarative description of the semantics of the behavior of a reasoning process; each reasoning trace can be seen as one of the alternatives for the behavior.

2.4. Reasoning by assumption

The specific reasoning pattern used in this article to illustrate the approach is reasoning by assumption. This type of reasoning often occurs in practical reasoning; for example, in

- Diagnostic reasoning based on causal knowledge
- Everyday reasoning
- Reasoning based on natural deduction

An example of diagnostic reasoning by assumption in the context of a car that will not start is

"Suppose the battery is empty, then the lights won't work. But if I try, the lights turn out to work. Therefore the battery is not empty."

Note that on the basis of the assumption that the battery is empty and causal knowledge that without a functioning battery the lights will not burn, a prediction is made on an observable world fact, namely, that the lights will not burn. After this an observation is initiated that has a result (lights do burn) that contradicts the prediction. Based on this outcome the assumption is evaluated and, as a result, rejected.

An example of an everyday process of reasoning by assumption is

"Suppose I do not take my umbrella with me. Then, if it starts raining at 5 p.m., I will get wet, which I don't want. Therefore I better take my umbrella with me."

Again, based on the assumption, some prediction is made, this time using probabilistic knowledge that it may rain at 5 p.m. The prediction is in conflict with the desire not to get wet. The assumption is evaluated and rejected.

Examples of reasoning by assumption in natural deduction are

- Reductio ad absurdum, or method of indirect proof
  After assuming A, I have derived a contradiction. Therefore I can derive not-A.
- Implication introduction
  After assuming A, I have derived B. Therefore I can derive that A implies B.
- Reasoning by cases
  After assuming A, I have derived C. Also after assuming B, I derived C. Therefore I can derive C from A or B.

Notice the common pattern in all of the examples presented: It seems that first a reasoning state is entered in which some fact is *assumed.* Next (possibly after some intermediate steps) a reasoning state is reached where *consequences* of this assumption have been *predicted.* Moreover, in some cases *observations* can be performed, obtaining additional information about the world to be included in a next reasoning state. Finally, a reasoning state is reached in which an *evaluation* has taken place, for example, resulting in rejection of the assumption; possibly in the next state the assumption actually is retracted, and further conclusions are added.

In Jonker and Treur (in press), this common pattern has been taken as a basis for the development of a simulation model for reasoning by assumption. According to this model, the process of reasoning by assumption involves three important subprocesses: *assumption determination, observation result prediction,* and *assumption evaluation.* See Fig. 1 for an overview of the model. In this figure, the rounded rectangles denote different components of the model where the different subprocesses take place (including the external world, which is used to observe the relevant predictions made). The arrows indicate information flow. Note that this model can be viewed as a refinement of Simon and Lea's (1974) dual-problem spaces model (see also, Klahr & Dunbar, 1988), which distinguishes between a space for generation of hypotheses and a space for evaluation of these hypotheses. In the model depicted in Fig. 1, an additional space is introduced for the prediction of the consequences of the hypotheses. In the original dual-problem spaces model, this space was considered to be part of the space for hypothesis generation.

In the remainder of this article, the previously mentioned model for reasoning dynamics is taken as a point of departure in the formal analysis of human reasoning traces.

3. A Temporal Trace Language

In recent literature on computer science and artificial intelligence, temporal languages to specify dynamic properties of processes have been put forward (e.g., Dardenne, van Lamsweerde, & Fickas, 1993; Dubois, Du Bois, & Zeippen, 1995; Herlea, Jonker, Treur, & Wijngaards, 1999). To specify properties on the dynamics of *reasoning* processes in particular, TTL (Herlea et al., 1999; Jonker & Treur, 1998) is adopted. This is a language in the family of languages to which situation calculus (Reiter, 2001) and event calculus (Kowalski and Sergot,

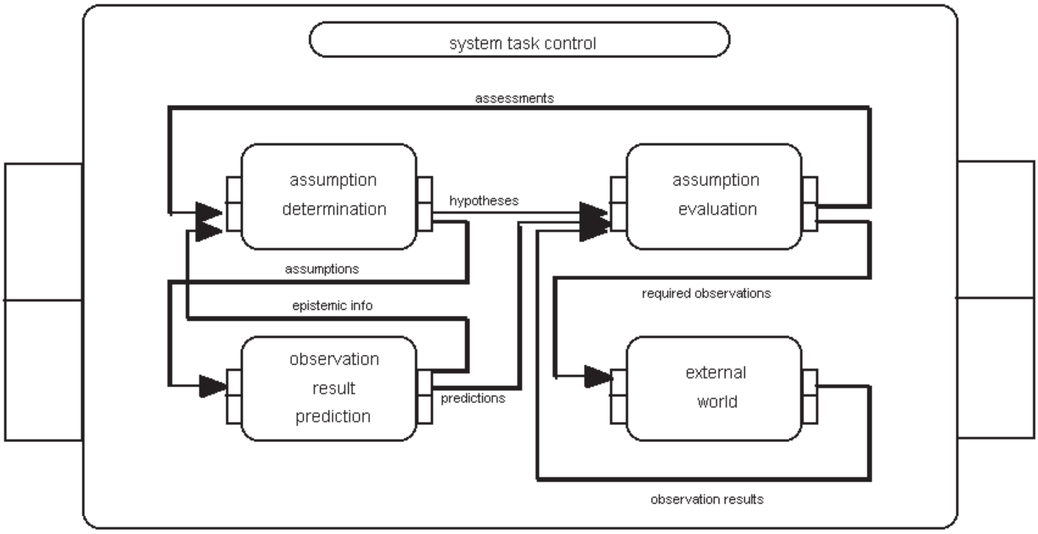


Fig. 1. Model for reasoning by assumption.

1986) also belong, and it was also successfully used to analyze multirepresentational reasoning processes in Jonker and Treur (2002).

3.1. Ontology

An *ontology* is a specification (in order-sorted logic) of a vocabulary. For the example reasoning pattern (i.e., reasoning by assumption in a game of Master Mind), the state ontology was inspired by the model depicted in Fig. 1 and includes unary relations such as assumed and rejected_code on sort ASSUMPTION, and binary relations such as prediction_for on RESULT × ASSUMPTION. The sort ASSUMPTION includes specific functions for domain statements such as code(COLOR, COLOR, COLOR). The complete ontology for this current domain is given in Table 1.

3.2. Reasoning state

A reasoning state for ontology Ont is an assignment of truth values {true, false} to the set of ground atoms At(Ont). The set of all possible states for ontology Ont is denoted by STATES(Ont). A part of the description of an example reasoning state S is:

assumed(code(red, white, blue))	: true
prediction_for(answer(black, empty, empty), code(red, white, blue))	: true
observation_result_for(answer(white), code(red, white, blue))	: true
rejected_code(code(red, white, blue))	: false

Table 1
Ontology for Reasoning in the Master Mind Domain

Relations and Sorts	Informal Description
Unary Relations	
focus_assumed(F:FOCUS)	The agent currently assumes F to be part of the solution.
assumed(A:ASSUMPTION)	The agent currently assumes A to be the solution.
rejected_code(A:ASSUMPTION)	The agent has rejected the assumption A.
rejected_focus(F:FOCUS)	The agent has rejected the focus assumption F.
Binary Relations	
prediction_for(R:RESULT, A:ASSUMPTION)	The agent predicts that if A is true, then R should be observable.
code_extension_for(A:ASSUMPTION, F:FOCUS)	The agent believes that if F is part of the solution, then the whole solution should be A.
to_be_observed_for(answer, A:ASSUMPTION)	The agent starts observing what the answer is for A.
observation_result_for(R:RESULT, A:ASSUMPTION)	The agent observes that R is the answer for the guess A.
Sorts	
FOCUS	at(C:COLOR, P:POSITION)
ASSUMPTION	code(C1:COLOR, C2:COLOR, C3:COLOR)
RESULT	answer(P1:PIN, P2:PIN, P3:PIN)
COLOR	{black, blue, brown, green, orange, red, white, yellow)
POSITION	{1, 2, 3}
PIN	{black, white, _}

RS is the sort of all reasoning states of the agent. For simplicity in the formulation of properties WS is the set of all substates of elements of RS, thus WS is the set of all world states. The standard satisfaction relation |== between states and state properties is used: S |== *p* means that state property *p* holds in state S. For example, in the reasoning state S previously mentioned, it holds S |== assumed(code(red, white, blue)).

3.3. Reasoning trace

To describe dynamics, explicit reference is made to time in a formal manner. A fixed time frame T is assumed that is linearly ordered. Depending on the application, it may be dense (e.g., the real numbers), or discrete (e.g., the set of integers or natural numbers or a finite initial segment of the natural numbers). A trace γ over an ontology Ont and time frame T is a mapping γ: T $\rightarrow$ STATES(Ont), that is, a sequence of reasoning states γ_t ($t \in$ T) in STATES(Ont). The set of all traces over ontology Ont is denoted by Γ(Ont), that is, Γ(Ont) = STATES(Ont)^T. The set Γ(Ont) is also denoted by Γ if no confusion is expected.

3.4. Expressing dynamic properties

States of a trace can be related to state properties via the formally defined satisfaction relation |== between states and formulas. Comparable to the approach in situation calculus, the sorted predicate logic TTL is built on atoms such as state(γ, t) |== p, referring to traces, time, and state properties. This expression denotes that state property p is true in the state of trace γ at time point t. Here |== is a predicate symbol in the language (in infix notation), comparable to the Holds-predicate in situation calculus. Temporal formulas are built using the usual logical connectives and quantification (e.g., over traces, time, and state properties). The set TFOR(Ont) is the set of all temporal formulas that only make use of ontology Ont. We allow additional language elements as abbreviations of formulas of the temporal trace language. The fact that this language is formal allows for precise specification of dynamic properties. Moreover, editors can and actually have been developed to support specification of properties. Specified properties can be checked automatically against example traces to find out whether they hold.

4. The experiment

To illustrate the benefits of the analysis approach mentioned previously, it has been applied in a case study. A simple reasoning experiment in the domain of Master Mind was performed to obtain empirical data about reasoning by assumption. This section describes the experiment in detail.

4.1. Participants

Thirty persons with different social background participated in the experiment. The group consisted of 19 men and 11 women. Their mean age was 28.2 years, with a standard deviation of 10.0.

4.2. Method

The participants were asked to solve a simplified game of Master Mind. Before starting the experiment, they were given the following instructions:

> The opponent picks a secret code consisting of three pegs, each peg being one of eight colors. Your goal is to guess the exact positions of the colors in the code in as few guesses as possible. After each guess, the opponent gives you a score of exact and partial matches. For each of the pegs in your guess that is the correct color in the correct position, the opponent will give you an “exact” point (represented by a black pin). If you score 3 black pins on a guess, you have guessed the code. For each of the pegs in the guess that is a correct color in an incorrect position, the opponent will give you an “other” point (represented by a white pin). Together, the black and white pins will add up to no more than 3. Notice that the positions of the black and white pins do not necessarily relate to the positions of the colors. Within this specific experiment, *one initial guess has already been done for you.* While doing the experiment, please think aloud, explaining each step you perform.

For each participant, the *solution code* was the same, namely the combination [blue–white–red]. The *initial guess* mentioned previously was always the combination [red–white–blue]. Hence, the provided answer corresponding to the initial guess was [black–white–white].

In Tables 2 and 3, two example traces are shown, including the way in which they were formalized to automatically check their properties. The left column contains the human transcript, the right column contains the formal counterpart. Two additional examples can be found in Appendix C. The transcripts of all human reasoning traces can be found at http://www.cs.vu.nl/~tbosse/mastermind/human-traces.doc

5. Dynamic properties

In this section a number of dynamic properties that have been identified as relevant for patterns of reasoning by assumption are presented. As mentioned in the Introduction, two categories of dynamic properties are distinguished. The first category is specified by *characterizing properties.* These are properties that are expected to hold for all reasoning traces. In contrast, the second category contains *discriminating properties,* properties that distinguish several types of traces from each other. Within each category, *global properties* (GPs, addressing the overall reasoning behavior) as well as *local properties* (LPs, addressing the step-by-step reasoning process) are given. Note that the properties are not given in any particular order and that their numbering has no special meaning.

5.1. Characterizing properties

Based on the model presented in Section 2.4, a number of *characterizing properties* have been expressed for the pattern of reasoning by assumption. These properties are shown as follows, both in an informal and a formal notation (in TTL).

GP1 termination of assumption determination
The generation of new assumptions will not go indefinitely.
∀γ:Γ ∃*t*:T ∀A:INFO_ELEMENT
∀*t′*:T ≥ *t*:T [state(γ,*t′*) |== assumed(A) ⇒ state(γ, *t*) |== assumed(A)]

This property holds for all traces, which is not surprising, because the experiments did not last forever.

GP2 correctness of rejection
Every code that has been rejected does not hold in the world situation.
∀γ:Γ ∀*t*:T ∀A:INFO_ELEMENT
state(γ,*t*) |== rejected_code(A) ⇒
state(γ,*t*) |=/= holds_in_world_for(answer(black, black, black), A)

This property holds for all traces, leading to the conclusion that none of the participants make the error of rejecting a code that is actually the solution. However, this does not necessarily imply that none of the participants rejects partial information. To find out whether this is the

Table 2
Example of Human Reasoning Trace

Human Transcript	Formalization
So, this is the first guess. Right. The national flag of Holland. *Exactly.*	
And this means that one of the colors is in the good place … and good color and good place, and also the other two colors are correct but they are not in the good place. *Exactly.*	
Right? Okay. So, what I'm going to do now. I'm going to … I'm trying to find out which of the colors is in a good place, first. So, let's say I say it's the red one. Maybe.	focus_assumed(at(red, 1))
So, I'm going to put the red here. And then, change these two.	code_extention_for(code(red, blue, white), at(red, 1)) assumed(code(red, blue, white)) prediction_for(answer(black, black, black), code(red, blue, white))
[red-blue-white] *Okay, so this is your guess?* This is my guess.	to_be_observed_for(answer, code(red, blue, white))
Then my answer is like this … [*white-white-white*] *… two, and three.*	observation_result_for(answer(white, white, white), code(red, blue, white))
Okay, so it wasn't the red. Okay.	rejected_code(code(red, blue, white)) rejected_focus(at(red, 1))
I will always use these ones, apparently. Then, keep the white and exchange red and blue.	focus_assumed(at(white, 2)) code_extention_for(code(blue, white, red), at(white, 2)) assumed(code(blue, white, red)) prediction_for(answer(black, black, black), code(blue, white, red))
[blue-white-red] *Okay, so why do you do this?* I'm testing now if the white one is in the good position.	to_be_observed_for(answer, code(blue, white, red))
Okay. So then my answer is this. Congratulations!	
[*black-black-black*]	observation_result_for(answer(black, black, black), code(blue, white, red))

case, an additional property should be needed, concentrating on rejected_focus instead of rejected_code.

GP3 completeness of rejection
After termination, all assumptions that do not hold in the world situation have been rejected.
$\forall\gamma{:}\Gamma\ \forall t{:}\mathrm{T}\ \forall\mathrm{A}{:}\mathrm{INFO\_ELEMENT}$
termination(γ,t)
$\wedge$ state(γ,t) |== assumed(A)
$\wedge$ state(γ,t) |=/= holds_in_world_for(answer(black, black, black), A)
$\Rightarrow$ state(γ,t) |== rejected_code(A)

Here termination(γ, t) is defined as $\forall\, t'{:}\mathrm{T}\ t' \geq t \Rightarrow \mathrm{state}(\gamma, t) = \mathrm{state}(\gamma, t')$.

Table 3
Example of Human Reasoning Trace

Human Transcript	Formalization
Well, at least the colors have already been determined.	
I want to know now … whether the red one was positioned correctly. *Okay.*	focus_assumed(at(red, 2))
[brown-red-__] Let's think now. Is this logical? I could of course also use one of the other colors twice. What would happen then? Then I can … Let's just see what happens then.	code_extension_for(code(blue, red, blue), at(red, 2)) assumed(code(blue, red, blue))
[blue-red-blue]	to_be_observed_for(answer, code(blue, red, blue))
These ones? Then the answer is as follows … [black-white]	observation_result_for(answer(black, white), code(blue, red, blue))
We know now that the white one was not placed correctly. No, we don't know that. Let's have a look, do we know that? No, we are not sure about that. It can also be that the blue one was in the right position, and that the white one was in the right position before that. Then it is the question whether this was a useful choice. At least it is the case that either … let's see now, if the red one was in the right position, then now the blue one is in the right position.	rejected_code(code(blue, red, blue))
But let me make the assumption that the white one was not in the right position.	focus_assumed(at(white, 1))
Then I would now … try this.	code_extension_for(code(white, red, blue), at(white, 1)) assumed(code(white, red, blue)) prediction_for(answer(black, black, black), code(white, red, blue))
[white-red-blue]	to_be_observed_for(answer, code(white, red, blue))
Then the answer is as follows … [*white-white-white*] *Like this.*	observation_result_for(answer(white, white, white), code(white, red, blue))
So nothing in the right position.	rejected_code(code(white, red, blue))
Let's see again. That means that in the first one the blue one was not in the right position either. So the blue one must be in the first or in the second.	rejected_focus(at(blue, 3))
The red one is not in the second …	rejected_focus(at(red, 2))
… so in the second only a blue one can have been right; that one is positioned in the first or in the second, so the blue one is on the first, that one is correct. So that we know already.	focus_assumed(at(blue, 1))
Furthermore, considering the white one. If the blue one should have been in the first, and we know that then … let's have a look, then there can be … then the white one has to be, according to that first one, to that first one it has to be correct.	code_extension_for(code(blue, white, red), at(blue, 1)) assumed(code(blue, white, red)) rejected_focus(at(white, 1))
Therefore this should be the solution.	prediction_for(answer(black, black, black), code(blue, white, red))
[blue-white-red]	to_be_observed_for(answer, code(blue, white, red))
All right. That is correct. [*black-black-black*] Good.	observation_result_for(answer(black, black, black), code(blue, white, red))

This property holds for all traces, implying that all participants eventually reject their incorrect assumptions. However, note that some of these rejections were made implicitly. For instance, consider the situation that a participant first assumes that the code is [red–blue–white], and subsequently assumes that the code is [blue–white–red]. In that case, the predicate rejected_code(red, blue, white) was included in the trace, but the participant did not state this explicitly.

GP4 guaranteed outcome
After termination, at least one evaluated assumption has not been rejected.
$\forall\gamma:\Gamma\ \forall t:T$
termination(γ,t) $\Rightarrow$
[$\exists$A:INFO_ELEMENT state(γ,t) |== assumed(A) $\wedge$ state(γ,t) |=/= rejected_code(A)]

This property holds for all traces, which indicates that every participant eventually finds the solution.

LP3 observation initiation effectiveness
For each prediction an observation will be made.
$\forall\gamma:\Gamma\ \forall t:T$ $\forall$A,B:INFO_ELEMENT
state(γ,t) |== prediction_for(B,A)
$\Rightarrow$ [$\exists t'$:T $\geq$ t:T state(γ,t') |== to_be_observed_for(answer, A)]

This property holds for all traces, leading to the conclusion that in every case that a prediction was made, this was followed by a corresponding observation.

LP4 Observation result effectiveness
If an observation is made the appropriate observation result will be received.
$\forall\gamma:\Gamma\ \forall t:T$ $\forall$A,B:INFO_ELEMENT
state(γ,t) |== to_be_observed_for(answer, A) $\wedge$
state(γ,t) |== holds_in_world_for(B,A)
$\Rightarrow$ [$\exists$t$'$:T $\geq$ t:T state(γ,t') |== observation_result_for(B,A)]

This property holds for all traces. Thus, in all traces, the opponent provided the correct answers.

LP5 Evaluation effectiveness
If an assumption was made and a related prediction is falsified by an observation result, then the assumption is rejected.
$\forall\gamma:\Gamma\ \forall t:T$ $\forall$A,B:INFO_ELEMENT
state(γ,t) |== assumed(A) $\wedge$ state(γ,t) |== prediction_for(B,A)
$\wedge$ state(γ,t) |== observation_result_for(C,A) $\wedge$ B $\neq$ C
$\Rightarrow$ [$\exists t'$:T $\geq$ t:T state(γ,t') |== rejected_code(A)]

This property, which relates to GP2, holds for all traces. Thus, all participants correctly rejected a certain assumption when they had reason to do this (i.e., when the corresponding prediction was falsified by an observation result).

5.2. Discriminating properties

An analysis in terms of characterizing properties as mentioned previously is useful to create and validate a generic theory on a specific type of reasoning. Here, by generic it is meant that the theory can be applied to any particular person who reasons by assumption, regardless of the specific strategy used. However, usually in reasoning tasks differences can also be observed between individuals. Therefore, it is useful to also specify a number of *discriminating properties* of reasoning by assumption. These properties are shown as follows, both in an informal and a formal notation. In addition, it mentions how many of the 30 participants hold in each property.

GP5 correctness of assumption
Everything that has been assumed holds in the world situation.
∀γ:Γ ∀*t*:T ∀A:INFO_ELEMENT
state(γ,*t*) |== assumed(A) ⇒
state(γ,*t*) |== holds_in_world_for(answer(black, black, black), A)

This property only holds in 4 of the 30 cases. By checking it, the participants who made only correct assumptions can be distinguished from those who made some incorrect assumptions during the experiment. Put differently, the participants who immediately make the right guess are distinguished from those who need more than one guess. The fact that only 4 of the 30 participants are successful in their first guess indicates (as could be expected) that there is no confounding in the experiment whereby the participants can pick up information about the correct guess.

GP6 assumption grounding
Everything that has been assumed was based on an underlying focus (and code extension).
∀γ:Γ ∀*t*:T ∀A:INFO_ELEMENT
state(γ,*t*) |== assumed(A)
⇒ [∃*t′*:T < *t*:T ∃B:INFO_ELEMENT
state(γ,*t′*) |== focus_assumed(B) ∧ state(γ,*t′*) |== code_extension_for(A,B)]

This property holds in 26 of the 30 cases. Hence, the majority of the participants always generate their assumptions in two steps: first, they assume a certain color for one of the three positions, and then they extend this focus with assumptions for the other two positions. In contrast, four cases were found where the participants did not reason this way. These participants assumed a certain code without an underlying focus. There are two possible explanations for this phenomenon. One is that they did in fact make the focus assumption internally, but this could not be derived with certainty from their externally observable behavior. The second explanation is that they did not quite understand the rules of the game and hoped to make some progress by simply choosing a random code.

GP7 observation effectiveness
For each assumption, the agent eventually obtains the appropriate observation result.
∀γ:Γ ∀*t*:T ∀A,B:INFO_ELEMENT
state(γ,*t*) |== assumed(A) ∧ state(γ,*t*) |== holds_in_world_for(B,A)
⇒ [∃*t′*:T ≥ *t*:T state(γ,*t′*) |== observation_result_for(B,A)]

This property states that the agent always obtains the appropriate observation result for a particular assumption. For example, if an assumption is completely correct in the world, then the appropriate observation result should be three times black. Thus, the property gives more information about the experimenter than about the participant. In the experiments, this property holds for all but three of the traces. In these three cases people make an assumption that cannot be right, according to the information they have. However, they correct themselves before they decide to observe the answer to this wrong assumption. Thus, the answer to the incorrect assumption is never obtained.

GP8 essential assumption
When a solution was found, this was due to the focus at(white, 2).
$\forall \gamma{:}\Gamma\ \forall t{:}\mathrm{T}$
$\quad \mathrm{termination}(\gamma,t) \wedge \mathrm{state}(\gamma,t) \mid\!\!== \mathrm{assumed(code(blue, white, red))}$
$\quad \Rightarrow [\ \exists t'{:}\mathrm{T} < \mathrm{t{:}T}$
$\qquad \mathrm{state}(\gamma,t') \mid\!\!== \mathrm{focus\_assumed(at(white, 2))} \wedge$
$\qquad \mathrm{state}(\gamma,t') \mid\!\!== \mathrm{code\_extension\_for(code(blue, white, red), at(white, 2))}]$

This property holds in 25 of the 30 cases. Thus, the majority of the participants found the solution, [blue–white–red], thanks to the assumption that the white pin was at Position 2. However, other strategies are used as well, for example, focusing on the red or the blue pin.

GP9 initial assumption
The first focus assumption made was at(red, 1).
$\forall \gamma{:}\Gamma\ \exists t{:}\mathrm{T}$
$\quad \mathrm{state}(\gamma,t) \mid\!\!== \mathrm{focus\_assumed(at(red, 1))}$
$\quad \wedge\ [\ \forall t'{:}\mathrm{T} < t{:}\mathrm{T}\ \forall \mathrm{A{:}INFO\_ELEMENT}$
$\qquad \mathrm{state}(\gamma,t') \mid\!\!== \mathrm{focus\_assumed(A)} \Rightarrow \mathrm{A = at(red, 1)}]$

This property holds in 18 of the 30 cases. Thus, 18 participants started reasoning by assuming that the red pin was at Position 1. There are two possible explanations for this overall preference. First, although it is stated in the experiment that the order of the evaluation pins has no meaning, some of the participants might be guided by this order (i.e., black–white–white) in the first guess. Second, some participants might have a preference to analyze the pins systematically from left to right, and therefore start by focusing on the red pin. Nevertheless, there were still 12 participants who started in a different way.

GP10 second assumption
The second focus assumption made was at(red, 2).
$\forall \gamma{:}\Gamma\ \exists t{:}\mathrm{T}$
$\quad \mathrm{second\_focus}(\gamma,t) \wedge$
$\quad \mathrm{state}(\gamma,t) \mid\!\!== \mathrm{focus\_assumed(at(red, 2))}$

Here second_focus(γ, t) is defined as
$\exists \mathrm{A{:}INFO\_ELEMENT}\ \mathrm{state}(\gamma,t) \mid\!\!== \mathrm{focus\_assumed(A)} \wedge$
$\exists t'{:}\mathrm{T} < t{:}\mathrm{T}\ \exists \mathrm{B{:}INFO\_ELEMENT}\ \mathrm{state}(\gamma,t') \mid\!\!== \mathrm{focus\_assumed(B)} \wedge$
$[\ \forall''{:}\mathrm{T} < t{:}\mathrm{T}\ \forall \mathrm{C{:}INFO\_ELEMENT}\ \mathrm{state}(\gamma,'') \mid\!\!== \mathrm{focus\_assumed(C)} \Rightarrow \mathrm{C = B}]$

This property holds in 3 of the 30 cases. This means that 3 participants based their second guess on the focus assumption at(red, 2). In fact, all of these 3 participants based their first guess on the focus assumption at(red, 1). Thus, in the first two guesses they consistently focused on the position of the red pin. This is an important property, because it distinguishes two different types of reasoners with respect to the second guess: those who keep their focus on red (but realize that it has to be in another position) versus those who shift to another color. Although both approaches eventually lead to the same solution, the difference is relevant, because the reasoning strategies used are clearly distinct.

LP2 prediction effectiveness
For each assumption that is made, a prediction will be made.
$\forall\gamma{:}\Gamma\ \forall t{:}T\ \forall A{:}INFO\_ELEMENT$
state(γ,t) |== assumed(A)
$\Rightarrow$ [$\exists t'{:}T \geq t{:}T\ \exists B{:}INFO\_ELEMENT$
state(γ,t') |== prediction_for(B,A)]

This property holds in 26 of the 30 cases. So in four cases the participants make an assumption for which no prediction is made. Three of these four traces have already been discussed at GP7. The fourth trace involves the situation of Table 3, where the participant uses the following reasoning pattern: "I could use one of the colors twice. What would happen then? Well, I don't know. Let's just see what happens. … " Hence, the participant tries a code that he or she intuitively thinks is an intelligent guess, without really understanding why. Therefore, he does not make a prediction.

LP2′ prediction optimism
For each assumption that is made, the prediction answer (black, black, black) will be made.
$\forall\gamma{:}\Gamma\ \forall t{:}T\ \forall A{:}INFO\_ELEMENT$
state(γ,t) |== assumed(A)
$\Rightarrow$ [$\exists t'{:}T \geq t{:}T$
state(γ,t') |== prediction_for(answer(black, black, black),A)]

This property is a variant of property LP2. It holds in 24 of the 30 cases. In these cases the participants predict, for every assumption they make, that it is the correct solution. Given the fact that the participants have no special talent for guessing (see GP5), it might be a bit surprising that so many of them still make guesses that they hope are correct, rather than using a more systematic strategy. See Section 8 for a more detailed discussion on this topic. Nevertheless, for 6 participants, property LP2′ does not hold. Four of these 6 participants are those who make no predictions at all (see LP2). The interesting cases, however, are the 2 participants who do make predictions, but predict that their assumptions are *not* entirely correct. It turned out that this way of reasoning was part of a deliberate strategy of the participants. What they did was make a focus assumption (e.g., a red pin is at Position 1), and then extending this focus by adding "neutral" colors (e.g., assuming the code [red–yellow-yellow]). By doing this, the participant already knows that his or her guess will not be entirely correct, but he or she still makes this guess to receive partial information of the solution in a very systematic way.

6. Results

A special piece of software has been developed that takes a formally specified property and a set of traces as input and verifies whether the property holds for the traces (see Bosse, Jonker, Schut, & Treur, 2004). By means of this checking software, all specified properties have been checked automatically against all traces to find out whether they hold. In Table 4 an overview of the results is shown. In this table, an X indicates that the property holds for that particular trace. The final row provides the number of guesses needed by each participant to solve the problem.

As can be seen in this table, all characterizing properties indeed hold for all traces. This contributes to the validation of the model presented in Section 2.4. However, note that this is only an empirical validation, based on a limited number of empirical traces.

As opposed to the characterizing properties, the discriminating properties only hold for some of the traces. Therefore, these results can be used to distinguish several types of transcripts from each other, thereby obtaining a classification of different reasoning strategies. To do this in a more structured way, some simple tree clustering techniques (Kaufman & Rousseeuw, 1990) have been used to reduce the number of different classes. To do this, the following procedure was used. In the first step, all rows indicating characterizing properties have been removed from the table, and the resulting individuals with the same properties have been clustered together. In the following steps, more rows have been removed from the table (in a stepwise manner, starting with the discriminating property that holds for most individuals, i.e., GP7). This process has been repeated until only four rows are left. The results can be seen in Table 5. These results suggest that most of the reasoners fall in the fourth class (for which the properties GP8, GP9, and LP2′ hold, and GP10 does not hold). This class of reasoners could be described as the "systematic reasoners," because their reasoning processes satisfy the following combination of properties:

- They start focusing on the left pin (GP9).
- They continue by focusing on another color that could have corresponded with the black pin (instead of focusing on red again, GP10).
- They only make guesses that are possible solutions (LP2′).
- They find a solution due to the focus on the white pin (GP8).

Another interesting class of reasoners is defined by all traces for which property LP2′ does not hold (i.e., a combination of Columns 2, 3, 5, and 7). This class of reasoners follows a rather specific strategy. They all start by using "wrong" colors to obtain information about part of the solution. Only after obtaining this partial information, do they start making guesses about the solution as a whole. Therefore, this class could be described as the *stepwise reasoners.* In a similar manner, some qualifications could be given to the other classes, such as *strategic reasoners* or *random reasoners.*

Table 4
Overview of the Experimental Results: Traces Against Properties

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
GP1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
GP2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
GP3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
GP4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
GP5	X	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	X	—	—	—	—	—	—	X	—	—	—	—
GP6	X	X	X	X	—	—	X	X	X	X	X	X	X	X	X	X	X	—	X	X	X	X	X	X	X	X	X	—	X	X
GP7	X	X	X	X	X	—	X	X	X	X	X	X	X	X	—	X	X	—	X	X	X	X	X	X	X	X	X	X	X	X
GP8	X	X	—	X	X	X	X	X	X	—	X	X	X	X	X	X	X	—	X	X	X	X	—	X	X	X	X	—	X	X
GP9	—	—	—	—	—	X	X	X	X	—	X	X	X	—	—	X	X	X	—	X	X	X	X	—	X	—	—	X	X	X
GP10	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	X	—	—	—	—	—	—	X	—	—	—	—	X	—	—
LP3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
LP4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
LP5	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
LP2	X	X	X	X	X	—	X	X	X	—	X	X	X	X	—	X	X	—	X	X	X	X	X	X	X	X	X	X	X	X
LP2´	X	X	—	X	X	—	X	X	X	—	X	X	X	X	—	X	X	—	X	X	—	X	X	X	X	X	X	X	X	X
steps	1	2	3	3	3	3	3	2	3	3	2	3	2	1	3	3	3	3	1	3	2	2	3	3	2	1	3	3	3	2

Table 5
Overview of the Results After Applying Tree Clustering

	1, 2, 4, 5, 14, 19, 24, 26, 27	3, 10	6, 21	7, 8, 9, 11, 12, 13, 17, 20, 22, 25, 29, 30	15	16	18	23, 28
GP8	X	—	X	X	X	X	—	—
GP9	—	—	X	X	—	X	X	X
GP10	—	—	—	—	—	X	—	X
LP2´	X	—	—	X	—	X	—	X

7. Logical relations

In addition, logical relations have been identified between properties at different abstraction levels. An overview of the identified logical relations relevant for overall property GP7 is depicted as an AND-tree in Fig. 2.

For example, the relation at the highest level expresses that IP0 & LP4 ⇒ GP7 holds. Here, IP0 is an *intermediate property,* expressing the dynamics of the reasoning between two milestones:

IP0 assumptions lead to observation initiation.
For each assumption that is made, a prediction will be made.
∀γ:Γ ∀*t*:T ∀A:INFO_ELEMENT
state(γ,*t*) |== assumed(A)
⇒ [∃*t*′:T ≥ *t*:T state(γ,*t*′) |== to_be_observed_for(answer, A)]

Intermediate properties address smaller steps than GPs do, but bigger steps than LPs do. At a lower level, Fig. 2 depicts the relation LP2 & LP3 ⇒ IP0.

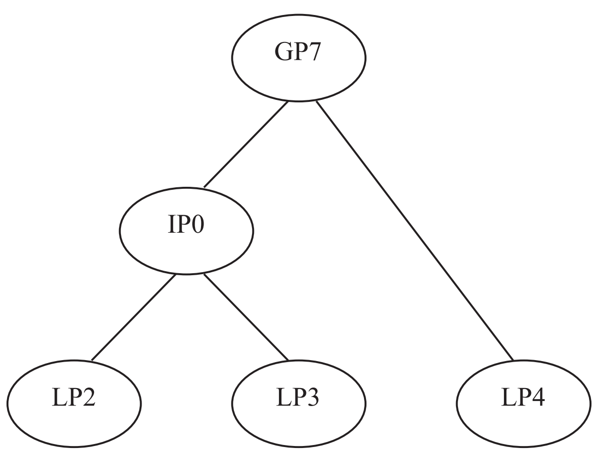


Fig. 2. Logical relations between dynamic properties.

Notice that the results given in Table 4 validate these logical relations. For instance, in all traces where LP2, LP3, and LP4 hold, GP7 also holds. Such logical relations between properties can be very useful in the analysis of empirical reasoning processes. For example, if a given person does not obtain the appropriate observation result for her assumption (i.e., property GP7 is not satisfied by the reasoning trace), then by a refutation process it can be concluded that either property IP0 or property LP4 fails (or both). If, after checking these properties, it turns out that IP0 does not hold, then either LP2 or LP3 does not hold. Thus, by this example refutation analysis it can be concluded that the cause of the unsatisfactory reasoning process can be found in either LP2 or LP3. In other words, either the *observation initiation* mechanism fails (LP3), or the *prediction* mechanism fails (LP2).

In this section, only one logical relation is shown. However, many more global, intermediate, and LPs for the pattern of reasoning by assumption, as well as the relations between them, can be found in Appendixes A and B.

8. Discussion

Within our experiment, the number of guesses needed by the participants to solve the problem varied between one and three. However, the participants that only needed one guess did not know beforehand that their guess would be correct. They were just lucky, because other solutions were possible, given the initial situation. Thus, their strategy was not optimal. Nevertheless, a number of studies exist that analyze optimal strategies in Master Mind; for example (Knuth, 1977; Koyama & Lai, 1993). With respect to our specific (simplified) problem, it turns out that there are optimal strategies that can always solve the problem in two guesses. To apply such a strategy, one should start with a code involving one of the initial colors twice. For instance, [red–red–blue]. Making this guess will provide enough information to solve the problem in the next guess. The reason for this is that, given the initial situation, only three solutions are possible, namely [red–blue–white], [blue–white–red] and [white–red–blue]. And for each of these possible solutions, the guess [red–red–blue] will receive a unique answer, that is [black–white], [white–white], and [black–black], respectively.

Given the previously mentioned pattern, a natural question is why none of the participants used this optimal strategy. A first possible reason is that it seems unnatural for humans to make a guess when they know beforehand that it will not be the correct solution. Starting in the way described previously would feel like wasting a guess. This might explain part of the results, but as discussed in Section 6, some of the participants (the stepwise reasoners) did deliberately use some wrong colors (although they failed to discover the optimal strategy). So the question remains why these stepwise reasoners did not find the optimal strategy. A potential answer is that it appears to be difficult (or at least, not very attractive from a workload perspective) for the participants to start by exhaustively generating all possible solutions. If they would do that, they would find out that the problem in question is probably simpler than they expected, involving only three possible solutions. However, the workload needed to find this out is relatively high compared to the gain (of just one step). Still, a small subset of the participants did generate all possible solutions, but even they did not come up with an optimal strategy. Therefore some other inhibitory factors may have played a role as well. Examples of such factors are time and social pressure (some participants might be embarrassed when spending too much time on a rather simple problem) and motivational factors (the participants were not informed that the optimal solution could be found in two steps, so they were not really encouraged to try harder).

A final factor that may have played a role in the strategy selection of the participants is the specific domain of Master Mind. Possibly, in other application domains of the pattern of reasoning by assumption, other strategies are preferred. For example, in the domain of diagnosis, the strategy of only making guesses that are expected to hold (see property LP2′) could be less attractive. In this domain, people may be more likely to make assumptions that are expected *not* to hold, thereby eliminating causes in a systematic way (rule-out strategy). More research is needed to determine the extent to which the results found in this article can be generalized to other applications of reasoning by assumption.

9. Conclusion

This article introduces a novel approach for the analysis of reasoning processes and explores the applicability of the approach for the pattern of reasoning by assumption in the domain of Master Mind. The analysis approach is based on the formalization of empirical reasoning traces, and the automated analysis of dynamic properties. A variety of dynamic properties have been specified, some of which are considered characteristic for the reasoning pattern reasoning by assumption, whereas some other properties can be used to discriminate between different approaches to the reasoning. For the Master Mind experiments undertaken, properties of the first, characteristic, type were based on the model from Jonker and Treur (in press). These properties indeed turned out to hold for the acquired reasoning traces, which contributes to the empirical validation of the model. Properties of the latter, discriminating type hold for some of the traces and do not hold for other traces: They define subsets of traces that collect similar reasoning approaches. These subsets can be viewed as different classes of reasoners, such as systematic reasoners, stepwise reasoners, strategic reasoners, and random reasoners. In these experiments, the biggest class of reasoners was the systematic class. These persons started by focusing on the left pin, continued by focusing on the white or blue pin, and eventually found the solution due to a focus on the white pin. Moreover, during the whole experiment they only made guesses that are possible solutions. Nevertheless, several other strategies were observed. An interesting class was the class of stepwise reasoners, who tried to obtain partial information in a stepwise manner. Future research is necessary to find out whether these results are specific for the game of Master Mind, or whether they can be generalized to other applications of reasoning by assumption.

In addition, it was explained how logical relations can be established between dynamic properties at different levels (e.g., global dynamic properties are connected to local dynamic properties, via intermediate properties). It was shown that such interlevel relations may play an important role in the analysis of empirical reasoning processes. More specifically, it was shown how a refutation process can be used to localize the exact cause of failure of GPs that are expected to hold.

In addition to empirical traces, the analysis approach presented in this article can be applied to traces generated by simulation models. Dynamic properties found relevant for human traces can be used to validate a simulation model, by generating a number of simulation runs and checking the dynamic properties for the resulting traces. This type of validation has been exploited to validate a simulation model for reasoning by assumption to solve the wise men puzzle in Jonker and Treur (in press). Moreover, in Bosse, Jonker, and Treur (2003) a similar analysis approach was used to validate a simulation model for controlled multirepresentational reasoning involving arithmetic, geometric, and material representations.

Besides cognitive science, the analysis method can be relevant for the area of knowledge engineering. The aim in knowledge engineering is to formally model complex reasoning tasks, such as design or diagnosis. This contributes to modeling, design, evaluation, maintenance, validation and verification, and reuse of models (Fensel & van Harmelen, 1994; Treur & Wetter, 1993). Some previous work in knowledge engineering in the domain of problem solving is reported by Brazier, Treur, Wijngaards, and Willems (1999). In their article, the relevant domain knowledge is obtained mainly by means of interviews with domain experts. This work can be viewed as complementary to their work, because here the relevant domain knowledge is obtained by means of explicit experiments with a large number of participants.

With respect to future research, an interesting direction would be to observe the participants' reasoning behavior over multiple trials. Important questions in this respect are whether participants are able to discover and learn certain strategies, and whether experienced puzzlers perform better than novices. To answer these kinds of questions, for future work it is planned to perform a learning experiment where participants have to solve multiple puzzles at different time points.

Another possibility for further research is to compare this work with the work by Stenning and van Lambalgen (in press). In that article, the authors show that an important issue in human reasoning is often whether a classical or defeasible logic is used, and (in the latter case) whether a credulous or a skeptical stance is adopted. Therefore, they propose an alternative distinction in modeling reasoning processes (next to the traditional distinction between syntactic and semantic approaches), that is, a distinction between reasoning *toward* an interpretation and reasoning *from* an interpretation. They demonstrate that this distinction is more appropriate to explain empirical findings in reasoning, such as the suppression effect (Byrne, 1989). It remains to be investigated how this distinction connects with this research. One difference between our Master Mind experiment and the type of tasks considered in Stenning and van Lambalgen (in press) is that in the latter the relevant external information is given in natural language (i.e., a number of sentences), whereas in the former it has a more "mathematical" format (i.e., six colored pins). Therefore, in the type of reasoning modeled in this article the process of interpretation is less present (there is less room for different interpretations), so that it involves mainly reasoning *from* a fixed interpretation. Nevertheless, even in the Master Mind example there is still some reasoning *to* an interpretation. To investigate in more detail the role of interpretation in reasoning by assumption, it would be interesting to change the setup of the experiments in such a way that more explicit interpretation is needed.

Note

1. See Section 9 for a detailed discussion about this topic.

Acknowledgments

The authors are grateful to James Greeno, Keith Stenning, and an anonymous referee for their valuable comments on an earlier version of this article.

References

Bosse, T., Jonker, C. M., Schut, M. C., & Treur, J. (2004). Modelling shared extended mind and collective representational content. In M. Bramer, F. Coenen, & T. Allen (Eds.), *Research and Development in Intelligent Systems XXI, Proceedings of AI–2004, the 24th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence* (pp. 19–32). New York: Springer Verlag.

Bosse, T., Jonker, C. M., & Treur, J. (2003). Simulation and analysis of controlled multi-representational reasoning processes. In F. Detje, D. Doerner, & H. Schaub (Eds.), *Proceedings of the Fifth International Conference on Cognitive Modelling, ICCM'03* (pp. 27–32). Bamberg, Germany: Universitats-Verlag Bamberg.

Braine, M. D. S., & O'Brien, D. P. (Eds.). (1998). *Mental logic.* London: Lawrence Erlbaum Associates, Inc.

Brazier, F. M. T., Treur, J., Wijngaards, N. J. E., & Willems, M. (1999). Temporal semantics of compositional task models and problem solving methods. *Data and Knowledge Engineering, 29,* 17–42.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition, 31,* 61–83.

Dardenne, A., van Lamsweerde, A., & Fickas, S. (1993). Goal-directed requirements acquisition. *Science in Computer Programming, 20,* 3–50.

Dubois, E., Du Bois, P., & Zeippen, J. M. (1995). A formal requirements engineering method for real-time, concurrent, & distributed systems. In *Proceedings of the Real-Time Systems Conference, RTS'95*. Paris.

Fensel, D., & van Harmelen, F. (1999). A comparison of languages which operationalize and formalize KADS models of expertise. *Knowledge Engineering Review, 9,* 105–146.

Herlea, D. E., Jonker, C. M., Treur, J., & Wijngaards, N. J. E. (1999). Specification of behavioural requirements within compositional multi-agent system design. *Lecture Notes in AI, 1647,* 8–27. Retrieved November 10, 2005, from http://www.iids.org/publications/MAAMAW99.behreqs.pdf

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, England: Cambridge University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Jonker, C. M., & Treur, J. (1998). Compositional verification of multi-agent systems: A formal analysis of pro-activeness and reactiveness. *Lecture Notes in Computer Science, 1536,* 350–380. Extended version in: *International Journal of Cooperative Information Systems, 11,* 2002, pp. 51–92.

Jonker, C. M., & Treur, J. (2002). Analysis of the dynamics of reasoning using multiple representations. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society, CogSci 2002* (pp. 512–517). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Jonker, C. M., & Treur, J. (2003). Modelling the dynamics of reasoning processes: Reasoning by assumption. *Cognitive Systems Research Journal, 4,* 119–136.

Kaufman, L., & Rousseeuw, P. (1990). Finding groups in data: An introduction to cluster analysis. New York: Wiley.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12,* 1–55.

Knuth, D. E. (1977). The computer as master mind. *Journal of Recreational Mathematics, 9,* 1–6.

Koyama, K., & Lai, T. W. (1993). An optimal mastermind strategy. *Journal of Recreational Mathematics, 25,* 251–265.

Kowalski, R., & Sergot, M. (1986). A logic-based calculus of events. New Generation Computing, 4, 67–95.

Nelson, T. (2000). *A Brief History of the Master Mind™ Board Game.* Retrieved November 10, 2005, from http://www.tnelson.demon.co.uk/mastermind/history.html

Reiter, R. (2001). *Knowledge in action: Logical foundations for specifying and implementing dynamical systems.* Cambridge, MA: MIT Press.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking.* Cambridge, MA: MIT Press.

Schroyens, W. J., Schaeken, W., & d'Ydewalle, G. (2001). A meta-analytic review of conditional reasoning by model and/or rule: Mental models theory revised (Psychological report No. 278). Leuven, Belgium: University of Leuven, Laboratory of Experimental Psychology.

Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. Gregg (Ed.), *Knowledge and cognition* (pp. 105–128). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Stenning, K., & van Lambalgen, M. (in press). A working memory model of relations between interpretation and reasoning. *Cognitive Science Journal.*

Treur, J., & Wetter, T. (Eds.). (1993). *Formal specification of complex reasoning systems.* Chichester, England: Horwood.

Yang, Y., & Bringsjord, S. (2001). Mental MetaLogic: A new paradigm in psychology of reasoning [Abstract]. In L. Chen & Y. Zhuo (Eds.), *Proceedings of the Third International Conference on Cognitive Science, ICCS 2001* (pp. 199–204). Beijing: USTC Press.

Yang, Y., & Johnson-Laird, P. N. (1999). A study of complex reasoning: The case GRE "logical" problems. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society* (pp. 767–771). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Appendix A. Dynamic Properties

This appendix contains a number of dynamic properties that are relevant for the pattern of reasoning by assumption. All of the GPs and a random selection of the intermediate properties have been validated against the traces mentioned in Section 6, using automated checks as described in that section. Note that in some cases the terminology used in these properties does not completely match the terminology used in the properties given in Section 5. The reason for this is that the properties given in Section 5 are domain specific: They apply to the domain of Master Mind only, whereas the properties given here apply to the pattern of reasoning by assumption in general. For example, a number of them have been checked against human and simulation traces in another case study involving reasoning by assumption: the wise men puzzle (Jonker & Treur, 2003).

World assumptions

WP1 World consistency

If something holds in the world, then its complement does not hold.

∀γ:Γ ∀*t*:T ∀A:INFO_ELEMENT ∀S1,S2:SIGN
  state(γ,*t*) |== holds_in_world(A,S1) ∧ S1 ≠ S2 ⇒ state(γ,*t*) |=/= holds_in_world(A,S2)

∀γ:Γ ∀*t*:T ∀A:INFO_ELEMENT ∀S1,S2:SIGN
  state(γ,*t*) |=/= holds_in_world(A,S1) ⇒ state(γ,*t*) |== holds_in_world(A,S2) ∧ S1 ≠ S2

Domain assumptions

DK1 Domain knowledge correctness
All domain knowledge about assumptions implying predictions is correct.
$\forall\gamma:\Gamma\ \forall t:T\ \forall$A,B:INFO_ELEMENT $\forall$S1,S2:SIGN
state(γ,t) |== holds_in_world(A,S1) $\wedge$ domain_implies(A,S1,B,S2)
$\Rightarrow$ state(γ,t) |== holds_in_world(B,S2)]
$\forall\gamma:\Gamma\ \forall t:T\ \forall$A,B:INFO_ELEMENT $\forall$S1,S2:SIGN
state(γ,t) |=/= holds_in_world(A,S1) $\wedge$ domain_implies(A,S1,B,S2)
$\Rightarrow$ state(γ,t) |=/= holds_in_world(B,S2)]

Local properties

LP1 Assumption initialization
Make a first assumption.
$\forall\gamma:\Gamma$ $\forall$A:INFO_ELEMENT $\forall$S:SIGN
initial_assumption(A,S)
$\Rightarrow$ [$\exists\ t$:T state(γ,t) |== assumed(A,S)]

LP2 Prediction effectiveness
For each assumption that is made, all relevant predictions are generated.
$\forall\gamma:\Gamma\ \forall t:T\ \forall$A,B:INFO_ELEMENT $\forall$S1,S2:SIGN
state(γ,t) |== assumed(A,S1) $\wedge$ domain_implies(A,S1,B,S2)
$\Rightarrow$ [$\exists t'$:T $\geq$ t:T state(γ,t') |== prediction_for(B,S2,A,S1)]

LP3 Observation initiation effectiveness
All predictions made will be observed.
$\forall\gamma:\Gamma\ \forall t:T\ \forall$A,B:INFO_ELEMENT $\forall$S1,S2:SIGN
state(γ,t) |== prediction_for(B,S2,A,S1)
$\Rightarrow$ [$\exists t'$:T $\geq$ t:T state(γ,t') |== to_be_observed(B)]

LP4 Observation result effectiveness
If an observation is made the appropriate observation result will be received.
$\forall\gamma:\Gamma\ \forall t:T$ $\forall$A:INFO_ELEMENT $\forall$S:SIGN
state(γ,t) |== to_be_observed(A) $\wedge$ state(γ,t) |== holds_in_world(A,S)
$\Rightarrow$ [$\exists t'$:T $\geq$ t:T state(γ,t') |== observation_result(A,S)]

LP5 Evaluation effectiveness
If an assumption was made and a related prediction is falsified by an observation result, then the assumption is rejected.
$\forall\gamma:\Gamma\ \forall t:T\ \forall$A,B:INFO_ELEMENT $\forall$S1,S2,S3:SIGN
state(γ,t) |== assumed(A,S1) $\wedge$ state(γ,t) |== prediction_for(B,S2,A,S1)
$\wedge$ state(γ,t) |== observation_result(B,S3) $\wedge$ S2 $\neq$ S3

$\Rightarrow$ [$\exists t'$:T $\geq$ t:T state(γ,t') |== rejected(A,S1)]

LP6 assumption effectiveness
If an assumption is rejected, and there is still an alternative assumption available, this will be assumed.
$\forall\gamma$:Γ $\forall t$:T $\forall$A,B:INFO_ELEMENT $\forall$S1,S2:SIGN
state(γ,t) |== assumed(A,S1)
$\wedge$ state(γ,t) |== rejected(A,S1)
$\wedge$ state(γ,t) |== alternative_for(B,S2,A,S1)
$\wedge$ state(γ,t) |=/= rejected(B,S2)
$\Rightarrow$ [$\exists t'$:T $\geq$ t:T state(γ,t') |=/= assumed(A,S1) $\wedge$ state(γ,t') |== assumed(B,S2)]

Global properties

GP1 Termination of assumption determination
The generation of new assumptions will not go indefinitely.
$\forall\gamma$:Γ $\exists t$:T $\forall$A:INFO_ELEMENT, $\forall$S:SIGN
$\forall t'$:T $\geq$ t:T [state(γ,t') |== assumed(A,S) $\Rightarrow$ state(γ,t) |== assumed(A,S)]

GP2 Correctness of rejection
Everything that has been rejected does not hold in the world situation.
$\forall\gamma$:Γ $\forall t$:T $\forall$A:INFO_ELEMENT $\forall$S:SIGN
state(γ,t) |== rejected(A,S)
$\Rightarrow$ state(γ,t) |=/= holds_in_world(A,S)

GP3 Completeness of rejection
After termination, all assumptions that do not hold in the world situation have been rejected.
$\forall\gamma$:Γ $\forall t$:T $\forall$A:INFO_ELEMENT, $\forall$S:SIGN
termination(γ,t)
$\wedge$ state(γ,t) |== assumed(A,S)
$\wedge$ state(γ,t) |=/= holds_in_world(A,S)
$\Rightarrow$ state(γ,t) |== rejected(A,S)

P Persistence
Atoms are persistent (either unconditional or conditional).
$\forall\gamma$:Γ $\forall t$:T $\forall$A:INFO_ELEMENT $\forall$S:SIGN
state(γ,t) |== holds_in_world(A,S)
$\Rightarrow$ [$\forall t'$:T $\geq$ t:T state(γ,t') |== holds_in_world(A,S)]
$\forall\gamma$:Γ $\forall t$:T $\forall$A:INFO_ELEMENT $\forall$S:SIGN
state(γ,t) |=/= holds_in_world(A,S)
$\Rightarrow$ [$\forall t'$:T $\geq$ t:T state(γ,t') |=/= holds_in_world(A,S)]
$\forall\gamma$:Γ $\forall t$:T $\forall$A:INFO_ELEMENT $\forall$S:SIGN
state(γ,t) |== rejected(A,S)
$\Rightarrow$ [$\forall t'$:T $\geq$ t:T state(γ,t') |== rejected(A,S)]

$\forall\gamma{:}\Gamma\ \forall t{:}T\ \forall A{:}\text{INFO\_ELEMENT}\ \forall S{:}\text{SIGN}$
$\quad \text{state}(\gamma,t)\ |==\ \text{observation\_result}(A,S)$
$\quad \Rightarrow [\ \forall t'{:}T \geq t{:}T\ \text{state}(\gamma,t')\ |==\ \text{observation\_result}(A,S)]$
$\forall\gamma{:}\Gamma\ \forall t,t',t''{:}T\ \forall A{:}\text{INFO\_ELEMENT}\ \forall S{:}\text{SIGN}$
$\quad t \leq t'' \wedge \text{state}(\gamma,t)\ |==\ \text{assumed}(A,S)$
$\quad \wedge [t \leq t' \leq t'' \Rightarrow \text{state}(\gamma,t')\ |=/=\ \text{rejected}(A,S)]$
$\quad \text{state}(\gamma,t'')\ |==\ \text{assumed}(A,S)$
$\forall\gamma{:}\Gamma\ \forall t,t',t''{:}T\ \forall A,B{:}\text{INFO\_ELEMENT}\ \forall S1,S2{:}\text{SIGN}$
$\quad t \leq t'' \wedge \text{state}(\gamma,t)\ |==\ \text{prediction\_for}(A,S1,B,S2)$
$\quad \wedge [\ t \leq t' \leq t'' \Rightarrow \text{state}(\gamma,t')\ |=/=\ \text{rejected}(B,S2)]$
$\quad \text{state}(\gamma,t'')\ |==\ \text{prediction\_for}(A,S1,B,S2)$

Intermediate properties

IP1 Assumption existence uniqueness (1)
An assumption is never assumed twice.
$\forall\gamma{:}\Gamma\ \forall t{:}T\ \forall t'{:}T > t{:}T\ \forall A{:}\text{INFO\_ELEMENT},\ \forall S{:}\text{SIGN}$
$\quad \text{state}(\gamma,t)\ |==\ \text{assumed}(A,S) \wedge \text{state}(\gamma,t')\ |=/=\ \text{assumed}(A,S)$
$\quad \Rightarrow [\ \forall t''{:}T > t'{:}T\ \text{state}(\gamma,t'')\ |=/=\ \text{assumed}(A,S)]$

IP2 Possible assumption finiteness
There is a finite number N of possible assumptions (denoted by pa(A)).
$\text{card}(\text{pa}, N) \equiv \exists A_1 \ldots A_N\ [\ \wedge \wedge_{i\neq j}\ A_i \neq A_j \wedge \text{pa}(A_i) \wedge \forall A\ [\ \text{pa}(A) \Rightarrow \vee \vee_k\ A = A_k]]$

IP3 Assumption grounding
Each assumption that is assumed is a possible assumption.
$\forall\gamma{:}\Gamma\ \forall t{:}T\ \forall A{:}\text{INFO\_ELEMENT}\ \forall S{:}\text{SIGN}$
$\quad \text{state}(\gamma,t)\ |==\ \text{assumed}(A,S) \Rightarrow \text{pa}(A,S)$

IP4 Assumption retraction implies rejection
If something is assumed first, and later not assumed anymore, then it has been rejected.
$\forall\gamma{:}\Gamma\ \forall t{:}T\ \forall t'{:}T > t{:}T\ \forall A{:}\text{INFO\_ELEMENT},\ \forall S{:}\text{SIGN}$
$\quad \text{state}(\gamma,t)\ |==\ \text{assumed}(A,S) \wedge \text{state}(\gamma,t')\ |=/=\ \text{assumed}(A,S)$
$\quad \Rightarrow \text{state}(\gamma,t')\ |==\ \text{rejected}(A,S)$

IP5 Assumption existence uniqueness (2)
If something is rejected, then it will never be assumed again.
$\forall\gamma{:}\Gamma\ \forall t{:}T\ \forall A{:}\text{INFO\_ELEMENT},\ \forall S{:}\text{SIGN}$
$\quad \text{state}(\gamma,t)\ |==\ \text{rejected}(A,S)$
$\quad \Rightarrow [\ \forall t'{:}T > t{:}T\ \text{state}(\gamma,t')\ |=/=\ \text{assumed}(A,S)]$

IP6 Proper rejection grounding
If an assumption is rejected, then earlier on there was a prediction for it that did not match the corresponding observation result.

∀γ:Γ ∀*t*:T ∀A:INFO_ELEMENT ∀S1:SIGN
state(γ,*t*) |== rejected(A,S1)
⇒ [∃*t*′:T ≤ *t*:T ∃B:INFO_ELEMENT ∃S2,S3:SIGN
state(γ,*t*′) |== prediction_for(B,S2,A,S1) ∧ state(γ,*t*′) |== observation_result(B,S3) ∧ S2 ≠ S3]

IP7 Prediction–observation discrepancy implies assumption incorrectness
If a prediction does not match the corresponding observation result, then the associated assumption does not hold in the world.
∀γ:Γ ∀*t*:T ∀A,B:INFO_ELEMENT ∀S1,S2,S3:SIGN
state(γ,*t*) |== prediction_for(B,S2,A,S1) ∧ state(γ,*t*) |== observation_result(B,S3) ∧ S2 ≠ S3
⇒ state(γ,*t*) |=/= holds_in_world(A,S1)

IP8 Observation result correctness
Observation results obtained from the world indeed hold in the world.
∀γ:Γ ∀*t*:T ∀A:INFO_ELEMENT ∀S:SIGN
state(γ,*t*) |== observation_result(A,S) ⇒ state(γ,*t*) |== holds_in_world(A,S)

IP9 An Incorrect prediction implies an incorrect assumption (1)
If a prediction does not match the facts from the world, then the associated assumption does not hold either.
∀γ:Γ ∀*t*:T ∀A,B:INFO_ELEMENT ∀S1,S2,S3:SIGN
state(γ,*t*) |== prediction_for(B,S2,A,S1) ∧ state(γ,*t*) |== holds_in_world(B,S3) ∧ S2 ≠ S3
⇒ state(γ,*t*) |=/= holds_in_world(A,S1)

IP10 Observation result grounding
If an observation has been obtained, then earlier on the corresponding fact held in the world.
∀γ:Γ ∀*t*:T ∀A:INFO_ELEMENT ∀S:SIGN
state(γ,*t*) |== observation_result(A,S) ⇒ [∃*t*′:T ≤ *t*:T state(γ,*t*′) |== holds_in_world(A,S)]

IP11 An incorrect prediction implies an incorrect assumption (2)
If a prediction does not hold in the world, then the associated assumption does not hold either.
∀γ:Γ ∀*t*:T ∀A,B:INFO_ELEMENT ∀S1,S2:SIGN
state(γ,*t*) |== prediction_for(B,S2,A,S1) ∧ state(γ,*t*) |=/= holds_in_world(B,S2)
⇒ state(γ,*t*) |=/= holds_in_world(A,S1)

IP12 Prediction correctness
If a prediction is made for an assumption that holds in the world, then the prediction also holds.
∀γ:Γ ∀*t*:T ∀A,B:INFO_ELEMENT ∀S1,S2:SIGN
state(γ,*t*) |== prediction_for(B,S2,A,S1) ∧ state(γ,*t*) |== holds_in_world(A,S1)
⇒ state(γ,*t*) |== holds_in_world(B,S2)

IP13 Rejection effectiveness
If an assumption has been made and it does not hold in the world state, then it will be rejected.
∀γ:Γ ∀*t*:T ∀A:INFO_ELEMENT, ∀S:SIGN

state(γ,t) |== assumed(A,S)
∧ state(γ,t) |=/= holds_in_world(A,S)
⇒ [∃t':T ≥ t:T state(γ,t') |== rejected(A,S)]

IP14 An incorrect assumption implies prediction–observation discrepancy
If an assumption is made and it does not hold in the world, then a prediction for that assumption will be made that does not match the corresponding observation result.
∀γ:Γ ∀t:T ∀A:INFO_ELEMENT, ∀S1:SIGN
state(γ,t) |== assumed(A,S1)
∧ state(γ,t) |=/= holds_in_world(A,S1)
⇒ [∃t':T ≥ t:T ∃B:INFO_ELEMENT, ∃S2,S3:SIGN
state(γ,t') |== prediction_for(B,S2,A,S1) ∧ state(γ,t') |== observation_result(B,S3) ∧ S2 ≠ S3]

IP15 An incorrect assumption implies an incorrect prediction (1)
If an assumption is made, and it does not hold in the world, then a prediction for that assumption will be made that does not match the corresponding facts from the world.
∀γ:Γ ∀t:T ∀A:INFO_ELEMENT, ∀S1:SIGN
state(γ,t) |== assumed(A,S1)
∧ state(γ,t) |=/= holds_in_world(A,S1)
⇒ [∃t':T ≥ t:T ∃B:INFO_ELEMENT, ∃S2,S3:SIGN
state(γ,t') |== prediction_for(B,S2,A,S1) ∧ state(γ,t') |== holds_in_world(B,S3) ∧ S2 ≠ S3]

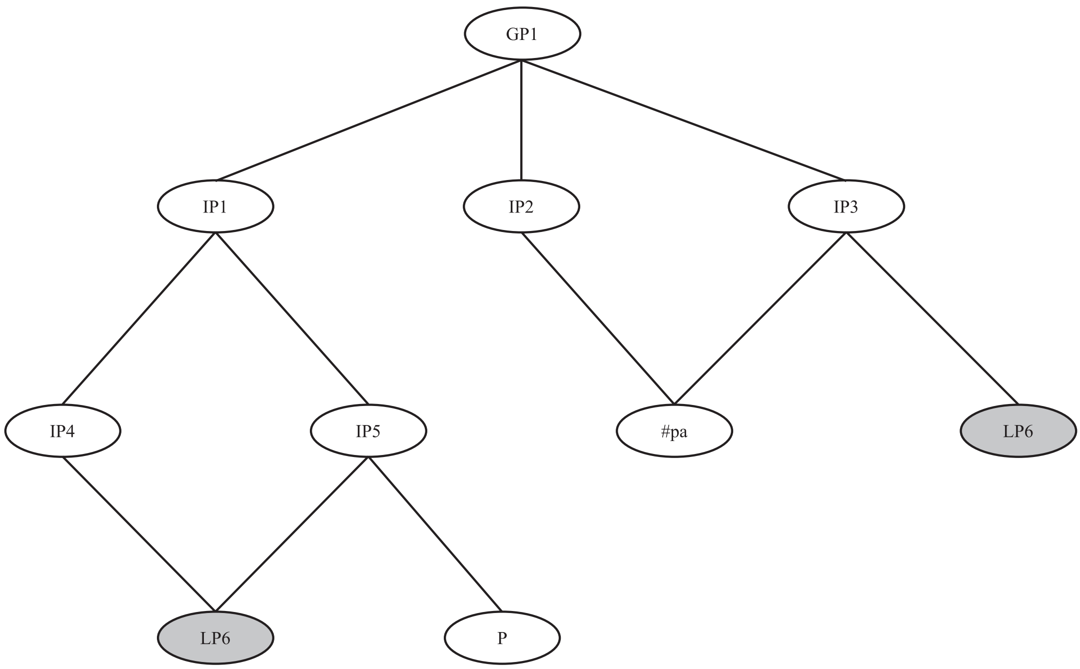
IP16 Observation effectiveness
For each prediction, the agent makes the appropriate observation.
∀γ:Γ ∀t:T ∀A,B:INFO_ELEMENT, ∀S1,S2,S3:SIGN
state(γ,t) |== prediction_for(B,S2,A,S1) ∧ state(γ,t) |== holds_in_world(B,S3)
⇒ [∃t':T ≥ t:T state(γ,t') |== observation_result(B,S3)]

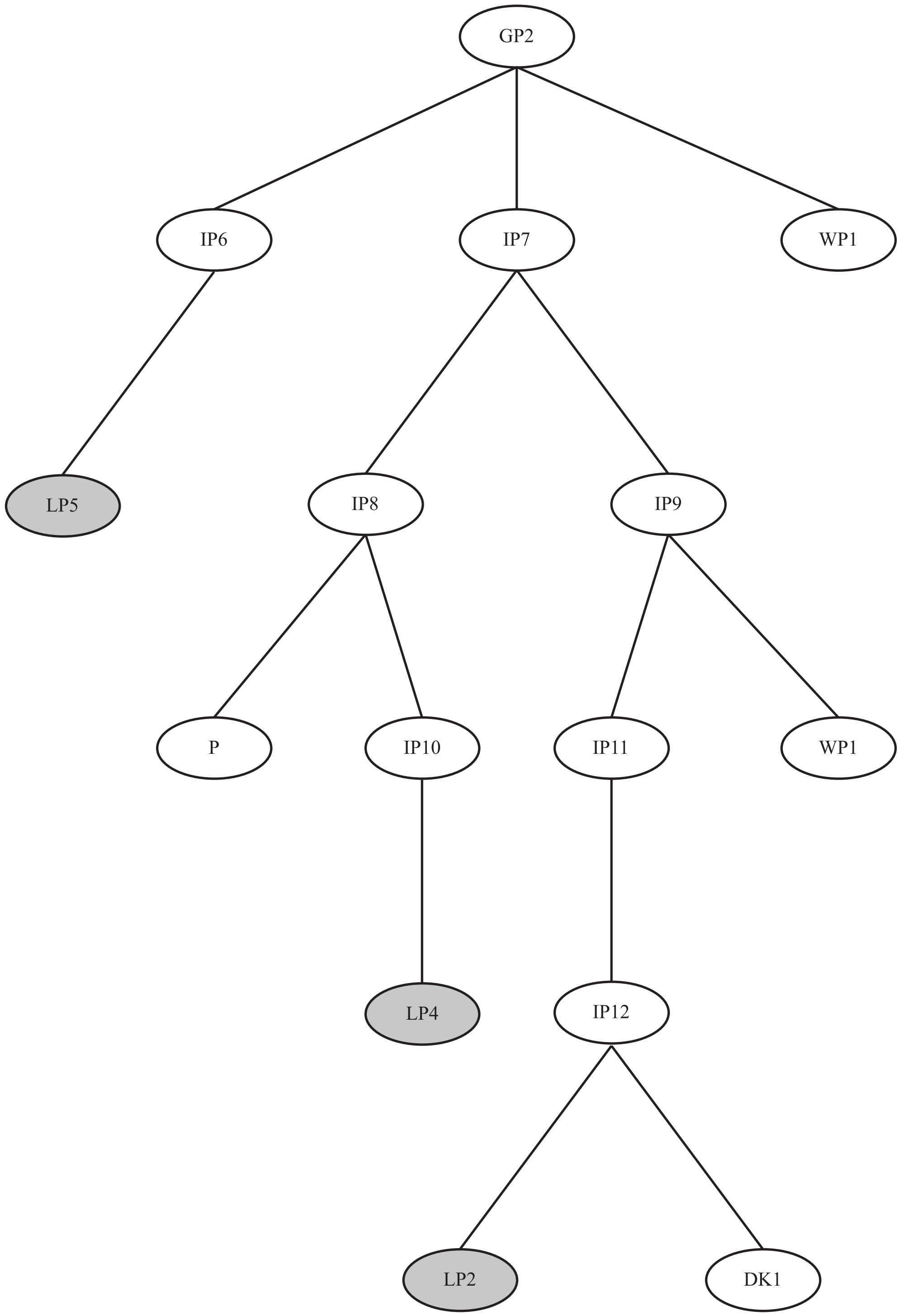
IP17 An incorrect assumption implies an incorrect prediction (2)
If an assumption is made, and it does not hold in the world, then a prediction for that assumption will be made that does not hold either.
∀γ:Γ ∀t:T ∀A:INFO_ELEMENT, ∀S1:SIGN
state(γ,t) |== assumed(A,S1)
∧ state(γ,t) |=/= holds_in_world(A,S1)
⇒ [∃t':T ≥ t:T ∃B:INFO_ELEMENT, ∃S2:SIGN
state(γ,t') |== prediction_for(B,S2,A,S1) ∧ state(γ,t') |=/= holds_in_world(B,S2)]

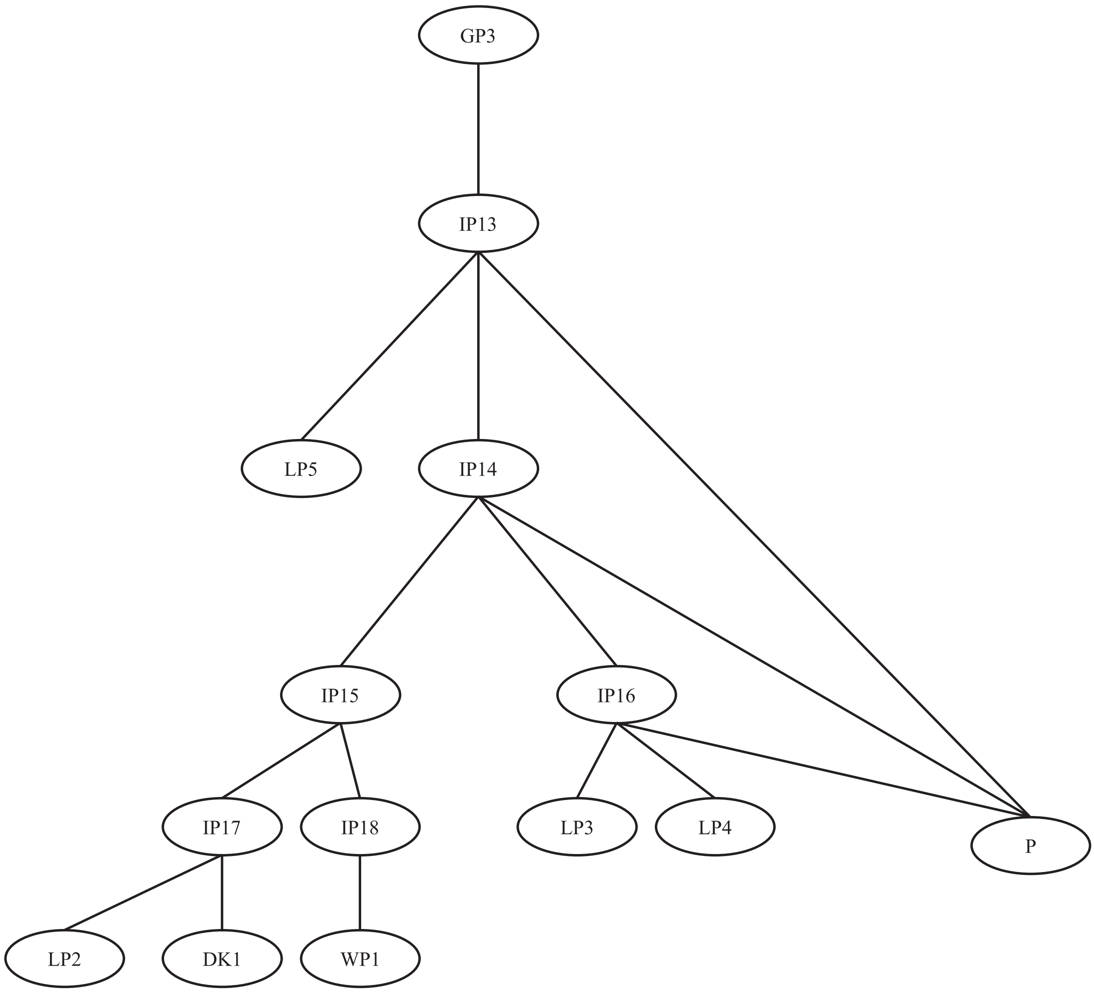
IP18 Prediction consistency
If a certain prediction does not hold in the world, then its complement does hold.
∀γ:Γ ∀t:T ∀A,B:INFO_ELEMENT, ∀S1,S2:SIGN
state(γ,t) |== prediction_for(B,S2,A,S1)
∧ state(γ,t) |=/= holds_in_world(B,S2)
⇒ [∃S3:SIGN state(γ,t) |== holds_in_world(B,S3) ∧ S2 ≠ S3]

Appendix B. Local Relations Between Dynamic Properties

This appendix contains a number of trees of logical relations relating global dynamic properties via intermediate dynamic properties to local dynamics properties. In particular, the following global dynamic properties have been worked out: GP1, GP2, and GP3. Here the gray ovals indicate that the "grounding" variant of the property is used, which states that the conclusion derived by that particular property is unique.




GP2
IP6
IP7
WP1
LP5
IP8
IP9
P
IP10
IP11
WP1
LP4
IP12
LP2
DK1



Appendix C. Transcripts

This appendix contains two additional transcripts, and their formalization. The left column contains the human transcript, the right column contains the formal counterpart. The complete set of transcripts of all human reasoning traces can be found at the following URL: http://www.cs.vu.nl/~tbosse/mastermind/human-traces.doc.

Example 1

Human transcript	Formalization
All right, so I will try to say aloud as much as possible. *Yes, please.* So a black one means that one is in the right position, and two white ones, that those are not in the right position. So all three colors are correct, because I already have three pins.	
Well, I just guess that white is in the right position …	focus_assumed(at(white, 2))
… and then I will swap the other two.	code_extension_for(code(blue, white, red), at(white, 2)) assumed(code(blue, white, red)) prediction_for(answer(black, black, black), code(blue, white, red))
[blue-white-red]	to_be_observed_for(answer, code(blue, white, red))
Okay. Why do you do this? Well, one of them is in the right position, so here I guessed one of them. And I know that these two colors are correct but that they are not in the right position so I have only those one as other possibility to change.	
Okay. Then my answer is very simple. That is already correct! [*black-black-black*]	observation_result_for(answer(black, black, black), code(blue, white, red))

Example 2

Human transcript	Formalization
Ooh! Well, all those three are already in. So that is easy. So all those others are not part of it. Well, let's have a look, a black one, so one of the three is correct and the other two I should swap. So then I can either just continue until I have it, or make use of others, that is also possible. What shall I do? I will make use of others, I like that. Like this.	
So that one is correct, that's what I think for the moment.	focus_assumed(at(red, 1))
And then I put two yellow ones in it. And then I will look what it becomes.	code_extension_for(code(red, yellow, yellow), at(red, 1)) assumed(code(red, yellow, yellow)) prediction_for(answer(black), code(red, yellow, yellow))
[red–yellow–yellow]	to_be_observed_for(answer, code(red, yellow, yellow))
Okay. Then the answer is like this … [*white*]	observation_result_for(answer(white), code(red, yellow, yellow))
Yes. So, I think then, the red one was not right in that position, so then it must have been one of the others.	rejected_code(code(red, yellow, yellow)) rejected_focus(at(red, 1))
So, now I will think, then it is for example the white one. That one was positioned correctly over there.	focus_assumed(at(white, 2))

(continued)

Example 2 *(Continued)*

Human transcript	Formalization
But the red one was not placed correctly over there, so then the red one should be over there. Let's have a look, is … am I doing that right? Perhaps I make it extra difficult for myself, and then it is still not correct. Let's have a look, well, let's try that anyway. Then it should be like this and then it should be like this …	code_extension_for(code(blue, white, red), at(white, 2)) assumed(code(blue, white, red)) prediction_for(answer(black, black, black), code(blue, white, red))
[blue–white–red]	to_be_observed_for(answer, code(blue, white, red))
Okay. Then the answer is like this …[*black–black–black*]	observation_result_for(answer(black, black, black), code(blue, white, red))
Yes!	
Congratulations!	