There Is No One Logic to Model Human Reasoning: the Case from Interpretation

Alexandra Varga[1], Keith Stenning[2], Laura Martignon[3]
[1] University of Giessen, Alexandra.Varga@psychol.uni-giessen.de
[2] University of Edinburgh, K.Stenning@ed.ac.uk
[3] Pädagogische Hochschule Ludwigsburg, martignon@ph-ludwigsburg.de

Abstract. The paper discusses a multiple-logics proposal for cognitive modelling of reasoning processes. It describes a staged view of human reasoning which takes interpretation seriously, and provides a non-technical introduction to a logic fit for modelling interpretative processes – Logic Programming. It summarises some results of the multiple-logics approach obtained with modelling psychological data, and with empirical tests of a combined use of reasoning strategies by human subjects. It draws some interim conclusions, and proposes avenues for future research.

1 Introduction

We are interested in computational models for human reasoning at the performance level. Cognitive modelling amounts to the use of some formalism in order to provide a productive description of cognitive phenomena. "Productive" has an explanatorily-oriented, twofold meaning: on the one hand, the description helps a better understanding of the phenomena, and second, it can be used to generate empirical predictions aiming to refine the theory that backs the model. By 'performance model' we imply that the formalism is actually used by real human agents in real reasoning contexts, wittingly or not. The reasoning process at the psychological level is an instantiation of the formal model. The 'wittingly or not' specification points to the need to include those forms of reasoning which are merely implicit, or below-awareness. A model of such reasoning processes involved in, e.g., understanding an utterance in one's native language, amounts to expressing these unwitting processes and subsequent behaviors 'as if' they were the result of computations expressed in a formal language.

We propose that the highest level of explanatory productivity, or information gain, can be achieved by a multiple-logics approach to cognitive modelling. In brief, this is so because of the complex differences between different kinds of reasoning which cannot be adequately captured by the formal properties of a single system. A multiple-logics approach is mandated because an all-purpose logic of human reasoning conflicts with the many things that humans may use reasoning for [1], e.g., to prove beyond reasonable doubt that the accused is guilty of the crime, to make the child understand the moral behind the story of the Ant and the Grasshopper. This would remain so even if all of the many formal candidates could be reconstructed in a single highly expressive logical system, because its use in human reasoning would be too resource-demanding; in other words, computational efficiency is an opportunity cost of expressive power. Performance models should at all points keep the balance.

Cognitive modelling from a multiple-logics perspective is also sanctioned by the history of psychological research. For instance, the withdrawal of previously validly derived conclusions when new information is added to the premise set [6], does not afford description in terms of a monotonic formalism such as classical logic. Everyday reasoning is most often non-monotonic. However monotonicity can be triggered by, e.g., by task instructions that create a dispute setting [1]. The bottom line is that different forms of reasoning, meant to achieve different goals, should be modelled in a formalism that bears the context-dependent properties of the inferences.

The main purpose of the current paper is to review the 'bridging potential' of a multiple-logics approach. The roadmap is as follows. We start in Section 2 by introducing the distinction between two kinds of reasoning, interpretation and further reasoning from that interpretation. We introduce the working example of a formalism, namely Logic Programming, and emphasise its application to interpretative processes. The remainder of the paper develops the argument based on taking interpretation seriously. Section 3 describes in detail a case of pre-linguistic implicit reasoning and summarises the modelling work in [35]. It shows how the logical and psychological aspects of reasoning can be integrated. Section 4 exemplifies the multiple-logics approach by describing the use of Logic Programming and fast and frugal heuristics for better understanding subjects' reasoning processes; we hemphasize the consequential methodological advantage of theoretical unification of the fields of reasoning and of judgement and decision-making. We end with some suggestions for further development of the multiple-logics approach, based on collaborative modelling among different systems.

2 The Proposed View of Reasoning and an Example of Formal Implementation

We are mostly concerned with everyday reasoning, i.e., the processes involved in habitual activities such as conversations, disputes, stories, demonstrations, etc. Stenning and van Lambalgen [32] set forth two kinds of processes: reasoning *to* an interpretation of the context, and reasoning *from* that interpretation.

Language processing is perhaps the clearest instantiation of the two reasoning stages. When speakers ask their interlocutors a question, they must first process the string of words in the context (linguistic and extra-linguistic) and produce an interpretation or model of it; in order to achieve the default purpose of communication fast and efficiently, these computations are aimed at the one model intended by the speakers. Because of this assumption that the right interpretation is in terms of what "(s)he must have meant to ask", the interpretative process is a paradigmatic case of credulous or cooperative reasoning. But this is only the beginning of the story. Should the first interpretation be unsatisfactory, e.g., being asked by one's life-time partner the question "How old are you?", hearers might resort to compensatory mechanisms, e.g., taking into account metaphorical meanings. Once a model is available the interlocutors can start to compute what they believe to be the contextually appropriate answer – this is reasoning *from* the interpretation. The reasoning path is not linear, e.g., additional utterances usually require model updates or re-computations of the initial discourse model.

The focus of cooperative interpretation on constructing a minimal contextual model can be described as the use of closed-world assumptions to frame the inferential scope [32, 34]. The basic format is the assumption for reasoning about abnormalities (CWA), which prescribes that, if there is no positive information that a given event must occur, one may assume it does not occur. These 'given events' are abnormalities with respect to the smooth, habitual running of a process; for example, a metaphorical interpretation is abnormal with respect to the literal one, and thus disregarded in minimal model construction. A conditional abnormality list is attached to each conditional; the list should be viewed as at the back of reasoners' minds [35]. That is, abnormalities are reasoned about only when evidence arrives (otherwise the assumption would be self-defeating). CWAs require construction of a minimal interpretation based only on what that is derivable from explicitly mentioned information. This is why they 'frame' [25] reasoning to manageable dimensions. Interpretation with CWAs is thus a plausible candidate to model the reasoning of agents with limited memory and computational resources in real-time.

The CWA is captured by all three parameters of Logic Programming – LP (syntactic, semantic, and definition of validity), a computational logic designed for automated planning [20]; it is the formal system that we use to instantiate our proposal. We view the utilization of such a formalism to model human inferences as a contribution to the bridge that this workshop seeks to build. Its cognitive plausibility has been shown from a variety of perspectives: it has been used to construct a formal semantics of tense [34], it helped understanding the formal structure of various cognitive tasks (e.g., Wason's task, the suppression task, the false belief task – dealt with in [32]), which in turn led to fine-grained experimental predictions (see [2] for a review).

Whereas an extensional formal approach deals with sets of items and with relations between those, an intensional one deals with characteristics and constitutive properties of the items in these classes. Relatedly, Logic Programming is an intensional formalism because its completion semantics is not directly truth-functional. We adopt the formal description of the logic set forth in [32, 34].

The CWA provides the notion of valid inference in LP, as truth preserving inferences in minimal models where nothing abnormal is the case. Relatedly, the LP conditional is represented as $p \;\&\; {\sim}ab \rightarrow q$ – "If *p* and nothing abnormal is the case, then *q*". Closed-world reasoning manifests itself in that, unless positive evidence (i.e., either explicit mentioning, or facts inferable from the database with the LP syntactic rules), the negation of the abnormality conjunct holds true. The syntactic expression of closed-world reasoning is the derivation rule of negation-as-failure – NAF. If a fact can be represented as the consequence of falsum $\bot$, thus it cannot be derived by backwards reasoning from program clauses, its negation is assumed true and the fact is thereby eliminated from the derivation. When resolving the query *q* given a program with clauses $p \;\&\; {\sim}ab \rightarrow q$ and $\bot \rightarrow ab$, *q* reduces to $p \;\&\; {\sim}ab$, from which *p* is derived by means of NAF. Use of negation-as-failure in derivations means that derivation checks if a query can be made true in a minimal model of the program. A minimal model is a 'closed world' in the sense that facts not forced to occur by inferences over the program clauses using the LP syntactic rules are assumed not to occur. The system's three-valued Kleene semantics (procedural in nature) warrants the construction of a unique minimal model, which is the only interpretation of concern of the current reasoning input[1]. Minimal models are provided by a semantic restriction of logic program clauses, called completion. It is obtained by introducing disjunction between all the bodies (antecedents) with the same head (consequent) in a program, and substituting implication with equivalence between the disjunctive body and the head.

The use of CWAs in interpretation is only the beginning of the intensional, or meaning-directed part of reasoning. Computations of a minimal preferred interpretation have been described at the psychological level in [32] as an interaction between the knowledge base of long-term memory and incoming input (e.g., new discourse statements, or new observations), in search for relevant information. Novel input may override the assumption and lead to subsequent model extensions by inclusion of the encountered abnormalities. This is a constitutively difficult task because at any give point, the vast majority of the long-term memory knowledge base is irrelevant. The Kleene semantics models this phenomenon by setting propositions to value U (undecided), which can develop to either T or F as a result of further inferences. The extensions of minimal models are also minimal. LP reasoning is thus inherently non-monotonic. Because of this it aligns with both the efficiency and the flexibility of everyday reasoning.

Let us relate this to the empirical sciences of human reasoning. What is most missing in the literature is detailed consideration of a positive account of the mental processes of interpretation, and of the interplay of the two forms of reasoning. In psychological experiments, when subjects are presented with the premises of a syllogism, they must first make sense of the information presented in order to be able to perform the inferences they are asked for. Reasoning to an interpretation must be acknowledged at face value by cognitive scientists when operationalizing theories into testable hypothesis, when deciding on the standards for response evaluation, when interpreting the empirical data, and obviously, when setting forth computational models for better understanding the cognitive phenomena. Despite a long period of utter neglect[2], recent work in the psychology of reasoning has started to acknowledge the role of interpretation, e.g., [18, 29]. This is a salutary new direction which calls for development of its consequences in modelling; consequently we argue that intensional formalisms are a necessary (though certainly not sufficient) ingredient of models for reasoning.

3 Logic for Modelling Implicit Reasoning

---

[1] Hölldobler and Kencana Ramli [15] criticised the Kleene semantics used in [35] by reference to modelling the suppression task [6]; these authors propose using the Lukasiewicz semantics instead. A technical rejoinder is available in the Appendix. Here we wish to emphasize that Byrne's task calls for a cooperative interpretation of the experimental material. The syntactic restrictions on LP conditionals on the other hand, e.g., non-iterability, allow completion to succeed in providing a minimal model as a pre-fixed point in a cooperative context, where epistemic trust is justified.

[2] A notable exception here is [17].

In a series of seminal studies with the head-touch task [12, 19], pre-linguistic infants have been shown to engage in selective imitative learning. We first introduce the experiment. After showing behavioral signs of being cold and wrapping a scarf around her shoulders, an adult demonstrates to 14-month-olds an unfamiliar head touch as a new means to activate a light-box. Half the infants see that the demonstrator's hands are occupied holding the scarf while executing the head action (Hands-Occupied condition – HO), the other half observe her acting with hands visibly free after having knotted the scarf (Hands-Free condition – HF). After a one-week delay subjects are given the chance to act upon the light-box themselves. They all attempt to light-up the lamp; however reenactment of the observed novel means action with the head is selective: 69% of the infants in the HF, and only 21% in the HO. More, [19] have shown that selectivity is contingent on a communicative action demonstration. This involves that throughout the demonstration session the experimenter behaves prosocially towards the infant, using both verbal and non-verbal communicative-referential cues. When the action was presented in a communicative context, the previous results were replicated. However, when the novel action is performed aloof, without infant-directed gaze or speech, the reenactment rate is always below chance level, and there is no significant difference between the HO and HF conditions. Gergely and his colleagues propose that infants' selectivity is underlain by a normative understanding of human actions with respect to goals. That is, infants learn some means actions but not others depending on the interpretation in terms of goals (teleological) afforded by the observed context.

The model set forth in [35] adopts this inferential perspective from the standpoint of multi-level teleology, i.e., a broad representation of goals that covers a whole range from physical goals (e.g., turning on a light-box) to higher-order intentions and meta-goals (e.g., the adult's teaching intention, infants' intentions to understand and to learn what is new and relevant)[3]. The inferential engine is constraint logic programming (CLP). The model gives voice to infants' interpretation of observations and to planning their own actions in the test phase. This voice is spelled out in the language of the event calculus [32] – 14-month-olds' observations and relevant bits of causal knowledge are represented as event calculus program clauses, e.g., *Initially*(communication) – agent exhibits infant-directed communicative behaviour, *Terminates*(contact, light-activity, t_k[4]) – contact is the culminating point of the light-box directed activity. Their teleological processing is called for and guided by the epistemic goals to understand and to learn, represented as integrity constraints [21, 34]. CLP allows to express higher-order goals as integrity constraints. These are peculiar conditional clauses which impose local (contextual) norms on the computations; they are universally quantified (but see footnote 6). For instance, IF ?*Initialy*(communication) succeeds THEN ?*HoldsAt*(teach$_f$, t) succeeds[5]

[3] Multi-level teleology is based on Kowalski's [21] distinction between achievement physical goals, and maintenance goals.

[4] t_k is a temporal constant.

[5] Note that the semantics of the conditional in integrity constraints is an unsettled issue [21]. [36] adopted a classical semantics.

expresses the assignment of a pedagogical intention to the observed agent conditional on her infant directed communicative behavior. When the antecedent is made true by the environment, i.e., in the communicative conditions, the young reasoner must act such that the goal expressed in the consequent becomes true. "teachf" is a parameterised fluent, i.e., a variable that must be specialized to a constant in the course of resolution. Infants' propensity for teleological understanding has been represented as an unconditional integrity constraint, namely ?*Happens*(x,t), *Initiates*(x,f(x),t), gx = f(x) succeeds. It demands assigning a concrete goal to an observed instrumental behaviour, i.e. finding a value for the Skolem function [6] *f(x)*. The requirement succeeds makes an existential claim with respect to a physical goal, i.e. there is such a state as *g*, which is a function *f(x)* of an action *x*.

Contextual interpretation amounts to finding the means – ends structure. Given the program clause *Initially*(communication) in the communicative condition, infants assign the adult the pedagogical intention expressed in the consequent of the constraint; further computations must unify parameter *f* with a concrete observed fluent, which is deemed to count as new and relevant information. Infants goal assignment to the agent's object-directed activity is done by resolving the unconditional constraint mentioned above. A successful unification is sought by specializing the function *f(x)* to a constant fluent from the narrative of events, given an evaluation of the causal relations available in the contextual causal model. The model shows how backward derivations from the constraint output the solution that the state *light-on* is the goal of contacting the light-box with the head, which is the culminating point of the observed activity. This represents infants' teleological conjecture, expected to render the action context understandable.

Interpretation is then subserved by a plan simulation algorithm – infants verify the goal conjecture by considering what they themselves would have done in order to achieve the goal *light-on*. This view of inferential plan simulation, and not merely motor simulation as traditionally construed, e.g., [26], is one of the main innovations brought about by this use of CLP for modelling. In the HO condition the mismatch between infants' closed-world plan calling for default hand contact, and observation of head contact is resolved by reasoning that the adult must use her hands for another goal, i.e., to hold the scarf in order not to be cold. The situation is fully understandable, hence infants specialize parameter *f* in ?*HoldsAt*(teachf ,t) to the object's newly inferred function, *light-on*.

The HO simulationist explanation does not work in the HF condition – the adult's free hands are not required to fulfill any different goal, so why it is that she does not use them to activate the object? Infants then integrate the adult's previously assigned pedagogical intentions in the explanatory attempt. Assigning a pedagogical intention to the reliable adult's otherwise incomprehensible head action renders it worth learning. Although touching a light box with the head in order to light it up may not be the most efficient action for the physical goal, the model proposes that it is considered efficient (and thereby reenacted) with respect to the adult's

[6] This is needed to handle the combination of universal and existential quantification – the existentially quantified variable within the scope of a universal quantifier is replaced with the value of a function of the universally quantified variable.

intention to share knowledge and the infant’s corresponding intention to learn.

In the test phase, upon re-encountering the light-box, infants plan their actions. The integrity constraint that guides their computations is ?*HoldsAt*($learn_f$,t), *Happens*(f’,t) succeeds; it corresponds to the adult’s pedagogical intention, and it expresses a ‘learning by doing’ kind of requirement. The outcome of interpretation, i.e., the means - ends structure of observations and the corresponding specialization of parameter *f*, modulate the constraint resolution. It sets up the physical goals that infants act upon in the test phase – either learn the new object’s function in HO (upon specialization of *f* to *light-on*), or also learn how to activate it in HF (upon specialization of *f* to $contact_{head}$). These goals are reduced to basic actions through the CLP resolution rule of backwards reasoning, which prescribes infants’ observed behaviour. In the HF condition thus, infants act upon two goals, learning the function and learning the means. The former goal is reduced to default hand actions (as required by closed-world reasoning), whereas the latter – to the novel head action. This explains infants’ performance of both hand and head actions. Reenactment of the head action can be described as ‘behavioural abduction’, a continuation in behavioural terms of the unsatisfactory explanatory reasoning.

The CLP model of observational imitative learning corroborates developmentalists’ argument that infants’ acquisition of practical knowledge from observation of adult agents is an instance of instrumental rationality. It does so by providing a concrete example of pre-linguistic reasoning to an interpretation, and of planning from the inferred means – ends structure of the situation. A logic is thus shown to be helpful in formalizing a quasi-automatic kind of reasoning, very different from the traditional understandings whereby playing chess, or proving mathematical theorems are the paradigmatic cases of reasoning. More research is needed in modelling other instances of fast and automatic reasoning processes, evidence of which is on the rise, e.g., [8].

4 A Joint Enterprise of Logic Programming and Heuristics for Reasoning and Decision-Making

We now show how a combined use of LP and its meta-analysis extension for counting can provide an account of causal reasoning. Martignon et al.’s [24] replication of Cummins’s [7] seminal results is an empirical proof that subjects’ judgments expressed in heuristic terms predict their confidence in conditional inferences. The authors propose that the use of fast and frugal heuristics is thus a method of reasoning to interpretations.

In the context of the ABC group, heuristics have inherited Einstein’s meaning [11]. That is, they are fast and frugal algorithms that “make us smart” *because* of their simplicity and not *in spite* of it [13]. In the field of judgement and decision-making they are specified as simple linear models for combining cues in tasks like comparison, estimation or categorization. There is extensive empirical evidence of their use, e.g., [5, 30]. Typical examples of heuristics are Take The Best – a linear model with non-compensatory weights, Tallying – a linear model with all weights equal to 1, or WADD – the weighted additive heuristic [27] whose weights are the cues validities (or ‘diagnosticities’).

Martignon et al. [24] set forth an analogy between the use of heuristics for combining cues in decision-making, and people's reasoning with defeaters. Consider for instance the causal conditional "If the brake was depressed then the car slowed down"; defeaters are cases when although the brake is depressed, the car does not slow down, e.g., the brake is broken. [9] showed that the more defeaters people generate, the less likely they are to endorse the conclusion of Modus Ponens. Martignon and colleagues recognized that it is precisely the Tallying heuristic on a profile of defeaters that is used for *combining* them in further inferences. This same heuristic is used for comparison decisions. In the typical comparison task analysed by [13], subjects must decide which of two German cities has a larger population, based on cues like "city *A* has a soccer team in the Bundesliga and city *B* does not", etc. When cues are abundant, subjects tend to tally them to make the comparison, and when cues are scarce, they rely on Take The Best, i.e., use the first cue that discriminates the cities and choose the one with the highest value [23].

So far cue ranking has been modeled in a Bayesian framework. Such ranking assumes that for each cue, e.g., having a soccer team in the Bundesliga, its validity is given by the probability that a city with a soccer team is larger than one without – a cue is valid when probability is larger than 0.5. This probabilistic computation has always been seen as cumbersome in the theory of fast and frugal heuristics [10], leading to serious doubts that probabilities can provide realistic performance models. LP on the other hand offers a simpler way for ranking cues. It is easy to see that a broken brake, for instance, can be represented as an abnormality in the LP representation of the conditional as $p \ \& \ {\sim}ab \rightarrow q$. The simpler way for ranking cues thus amounts to counting abnormalities for the conditional "If city *A* has a soccer team in the Bundesliga and city *B* does not, then city *A* is larger than city *B*". Here defeaters tallying will provide a good approximation of the conditional validity without complex probabilistic computations. In a similar vein, [24] have showed that other heuristics, like Best Cue [16] or WADD effectively predict subjects' confidence in the causal strength of the conditional. The crucial message is that LP can solve one aspect of modelling the use of heuristics in decision-making that has been criticized by other authors, namely relying on a Bayesian computation of cue validities [10]. This is so because LP facilitates heuristic selection compared with previously proposed modelling frameworks [22]. Ultimately, Martignon and colleagues [24] argue that LP may give a computational model of how the interpretations necessary for further probabilistic reasoning are arrived at.

It is a fascinating result that precisely the same heuristics that function so well for cue combination in judgment and decision-making are excellent for defeater combination in conditional reasoning. Because LP can easily model an interpretation of causal conditionals taking into account defeaters, and of the conditional expression of typical cues for decision-making, it provides a unified framework for the fields of (causal) reasoning, and of judgment and decision-making. This aligns with recent similar 'unificationist' approaches in the new paradigm of psychology of reasoning, e.g., [4].

5 Conclusions: Wrapping-up and Further-on

Despite the fact that gaps such as the one that gives the theme of the workshop are not easy to see in the raw data of the psychology of reasoning lab, to begin with however, their possibility must be acknowledged in order to allow for bridging. We started by presenting interpretation as an intrinsic, *sine qua non* stage of reasoning; this acknowledgement constrains realistic modelling endeavours to take it into account. We reviewed evidence that an approach to modelling which does take intensionality seriously by use of an expressive yet simple (at most linear on the name of nodes) formalism contributes to the theoretical integration of reasoning with judgement and decision-making. We also presented a computational model of pre-linguistic reasoning based on data from developmental psychology, and mentioned some consequences of this result for the ongoing debate with respect to dual-process theories of cognition.

With respect to future prospects for modelling applications of Logic Programming, we highlight the need for hypotheses of different domains where interpretation via minimal model construction may be adequate, and model that in terms of formalisms with minimal model semantics. The methodological implication of the multiple-logics proposal is a research program where modellers, given the properties of a particular formalism, hypothesise what kind of reasoning task it might model, and collaborate with experimenters to test those predictions; or observe properties of a reasoning task, hypothesise an appropriate formalisation, and test its empirical generalisations. With respect to LP, for instance, we propose that minimal model construction accurately models people's cooperative interpretation of conditionals uttered in a conversation setting [36]; investigations concerning other cases of cooperative reasoning, e.g., joint planning, joint intentionality, are current work in progress.

Throughout the paper we used LP to instantiate the multiple-logic proposal. Some other examples of applying non-deductive logics to human reasoning are Diderik Batens's program of adaptive logics [3], or Fariba Sadri's review of work on intention recognition [31]. It is noteworthy that both are essentially multiple-logic approaches. Consequently, last and most importantly, we wish to encourage pursuit of a multiple-system approach in research concerned with human reasoning. Our concrete suggestion concerns research on combining a logic that might appropriately model interpretation under computational constraints, i.e., in realistic cases of reasoning, with other formalisms such as probability [9]. One envisaged result is an alleviation of the problem of the priors, e.g., [28], by means of an intensional perspective offered by logics of interpretation. Such endeavour would bridge the gap between logical and AI systems for engineered reasoning, on the one hand, and empirical human reasoning research.

Appendix

In Chapter 7 of Stenning and van Lambalgen's *Human Reasoning and Cognitive Science* definite logic programs are used to represent non-monotonic reasoning with conditionals. The main technical tool is the interpretation of conditionals via the immediate consequence operator: the semantics is procedural, not declarative. This is because in a cooperative setting the truth of a conditional is not an issue, only what can be inferred from the conditional. This has consequences for what is meant by 'model of a program'. One may interpret the '$\to$' in program clauses truth-functionally, and say that $\mathcal{M} \models_3 \varphi \to q$ (where $\mathcal{M}$ is a 3-valued model) if the truth value of $\varphi \to q$ equals 1. Truth-functionality is not appropriate, since it would license nested occurrences of '$\to$', whereas nesting is not allowed by the syntax of logic programs, and hardly ever occur in natural language. Furthermore in this setting conditionals are never false, but apparent counterexamples are absorbed as 'abnormalities'. It follows that the expression 'model of a program P' cannot be given its literal meaning; its different sense is outlined below.

Let us start with the simpler case of positive programs. Recall that a positive logic program has clauses of the form $p_1 \wedge \ldots \wedge p_n \to q$, where the p_i, q are proposition letters and the antecedent (also called the body of the clause) may be empty. Models of a positive logic program P are given by the fixed points of a monotone operator:

Definition 1. *The operator T_P associated to a positive logic program P transforms a valuation $\mathcal{M}$ (viewed as a function $\mathcal{M} : L \longrightarrow \{0, 1\}$, where L is the set of proposition letters) into a model $T_P(\mathcal{M})$ according to the following stipulations: if v is a proposition letter,*

1. $T_P(\mathcal{M})(v) = 1$ *if there exists a set of proposition letters C, true on $\mathcal{M}$, such that* $\bigwedge C \to v \in P$
2. $T_P(\mathcal{M})(v) = 0$ *otherwise.*

Definition 2. *An ordering $\subseteq$ on (two-valued) models is given by: $\mathcal{M} \subseteq \mathcal{N}$ if all proposition letters true in $\mathcal{M}$ are true in $\mathcal{N}$.*

Lemma 1. *If P is a positive logic program, T_P is monotone in the sense that $\mathcal{M} \subseteq \mathcal{N}$ implies $T_P(\mathcal{M}) \subseteq T_P(\mathcal{N})$.*

Now consider the completion *comp*(P).

Definition 3. *Let $\mathcal{M}$ be a valuation. $\mathcal{M}$ is a* model of P *if $\mathcal{M} \models$ comp(P).*

Again it is easy to see that program clauses are not interpreted as truth functional implications, but rather as closure conditions on a model. This idea is best expressed using the operator T_P.

Lemma 2. *Suppose $\mathcal{M} \models$ comp(P). Then $T_P(\mathcal{M}) \subseteq \mathcal{M}$.*

PROOF. Application of T_P results in changing the truth value of atoms for which there is no immediate ground in the program P from 1 to 0. □

Definition 4. *A model* $\mathcal{M}$ *such that* $T_P(\mathcal{M}) \subseteq \mathcal{M}$ *is called a* pre-fixpoint *of* T_P*. It is* fixpoint *if* $T_P(\mathcal{M}) = \mathcal{M}$*.*

Let us next investigate the relation between completion, pre-fixpoints and fixpoints.

Lemma 3. *(Knaster-Tarski) A monotone operator defined on a directed complete partial order with bottom element (dcpo) has a least fixed point.*

In the simple situation considered (no negation), a model of the completion is a fixpoint of T_P and conversely, but this will no longer be true once negation is taken into account. Models of the completion *comp*(P) figure mostly when studying semantic consequences of the program P, therefore the following theorem provides all one needs:

Theorem 1. *Let* P *be a positive program, then there exists a fixpoint* $T_P(\mathcal{M}) = \mathcal{M}$ *such that for every positive formula*[1] F*:*

$$comp(P) \models F \Longleftrightarrow \mathcal{M} \models F.$$

PROOF. $\Leftarrow$ Choose a model $\mathcal{K} \models comp(P)$. The set of models $\{\mathcal{B} \mid \mathcal{B} \leq \mathcal{K}\}$ is a dcpo, hence T_P has a least fixed point $\mathcal{M} \subseteq \mathcal{K}$ here. Indeed, if **0** denotes the bottom element of the dcpo, then $\mathbf{0} \subseteq \mathcal{K}$ implies $T_P(\mathbf{0}) \subseteq T_P(\mathcal{K}) \subseteq \mathcal{K}$, whence it follows that the least fixpoint of T_P is a submodel of any $\mathcal{K} \models comp(P)$. By hypothesis $\mathcal{M} \models F$. Since F is positive and $\mathcal{M} \subseteq \mathcal{K}$, $\mathcal{K} \models F$, whence $comp(P) \models F$.
$\Rightarrow$ Since $\mathcal{M}$ is the least fixpoint of T_P, $\mathcal{M} \models comp(P)$, whence $\mathcal{M} \models F$. □

Definition 5. *A model* $\mathcal{K} \models comp(P)$ *is called* minimal *if there is no* $\mathcal{N}$ *which is a proper submodel of* $\mathcal{K}$ *(i.e. makes fewer atoms true).*

Lemma 4. *The least fixpoint of* T_P *is the unique minimal model of* $comp(P)$*.*

PROOF. Let $\mathcal{M}$ be the least fixpoint of T_P (which is obviously minimal). Let $\mathcal{K} \models comp(P)$ be another minimal model. Then since the bottom element $\mathbf{0} \subseteq \mathcal{K}$ and hence $T_P(\mathbf{0}) \subseteq T_P(\mathcal{K}) \subseteq \mathcal{K}$, it follows that $\mathcal{M} \subseteq \mathcal{K}$, which by minimality implies $\mathcal{M} \subseteq \mathcal{K}$. □

A 'minimal model of the program P' actually refers to the minimal model of the completion of P. Again, the difference is that to specify a model for P, one would need a declarative semantics for the arrow of logic programming, whereas no such thing is required in defining a model for the completion of P.

The needed logic programs must allow negation in the body of a clause, since the natural language conditional 'p implies q' is represented by the clause $p \wedge \neg ab \rightarrow q$. As observed above, extending the definition of the operator T_P with the classical definition of negation would destroy its monotonicity, necessary for the incremental approach to the least fixpoint. The pursued solution is to replace the classical two-valued logic by Kleene's strong three-valued logic, for which see figure 2.2. in Chapter 2. The equivalence $\leftrightarrow$ is defined by assigning 1 to $\varphi \leftrightarrow \psi$ if φ,ψ have the same truth value (in $\{u, 0, 1\}$) , and 0 otherwise.

We show how to construct models for definite programs, as fixed points of a three-valued consequence operator $\mathcal{T}_P^3$. We will drop the superscript when there is no danger of confusing it with its two-valued relative defined above.

[1] A formula containing only $\vee$, $\wedge$.

Definition 6. *A three-valued model is an assignment of the truth values* $u, 0, 1$ *to the set of proposition letters. If the assignment does not use the value* u*, the model is called* two-valued*. If* $\mathcal{M}, \mathcal{N}$ *are models, the relation* $\mathcal{M} \leq \mathcal{N}$ *means that the truth value of a proposition letter* p *in* $\mathcal{M}$ *is less than or equal to the truth value of* p *in* $\mathcal{N}$ *in the canonical ordering on* $u, 0, 1$.

Lemma 5. *Let* F *a formula not containing* $\leftrightarrow$*, with connectives interpreted using strong Kleene 3-valued logic; in particular* $\rightarrow$ *is defined using* $\neg$ *and* $\vee$*. Let* $\mathcal{M} \leq \mathcal{N}$*, then* $truth_{\mathcal{M}}(F) \leq truth_{\mathcal{N}}(F)$.

Definition 7. *Let* P *be a program.*

a. The operator $\mathcal{T}_P$ *applied to formulas constructed using only* $\neg$, $\wedge$ *and* $\vee$ *is determined by the strong Kleene truth tables.*
b. Given a three-valued model $\mathcal{M}$, $T_P(\mathcal{M})$ *is the model determined by*
   (a) $T_P(\mathcal{M})(q) = 1$ *iff there is a clause* $\varphi \rightarrow q$ *such that* $\mathcal{M} \models \varphi$
   (b) $T_P(\mathcal{M})(q) = 0$ *iff there is a clause* $\varphi \rightarrow q$ *in* P *and for all such clauses,* $\mathcal{M} \models \neg\varphi$
   (c) $T_P(\mathcal{M})(q) = u$ *otherwise*

The preceding definition ensures that unrestricted negation as failure applies only to proposition letters q which occur in a formula $\bot \rightarrow q$; other proposition letters about which there is no information at all may remain undecided. This will be useful later, when the operation of negation as failure is applied restrictively to ab only. Once a literal has been assigned value 0 or 1 by $\mathcal{T}_P^3$, it retains that value at all stages of the construction; if it has been assigned value u, that value may mutate into 0 or 1 at a later stage.

Lemma 6. *If* P *is a definite logic program,* T_P *is monotone in the sense that* $\mathcal{M} \leq \mathcal{N}$ *implies* $T_P(\mathcal{M}) \leq T_P(\mathcal{N})$.

Lemma 7. *Let* P *be a definite program.*

1. *The operator* $\mathcal{T}_P^3$ *has a least fixpoint, obtained by starting from the model* $\mathcal{M}_0$ *in which all proposition letters have the value* u*. By abuse of language, the least fixpoint of* $\mathcal{T}_P^3$ *will be called the minimal model of* P.
2. *There exists a fixpoint* $T_P^3(\mathcal{M}) = \mathcal{M}$ *such that for every formula* F *not containing* $\leftrightarrow$*:*
$$comp(P) \models F \Longleftrightarrow \mathcal{M} \models F;$$
   for $\mathcal{M}$ *we may take the least fixpoint of* $\mathcal{T}_P^3$.

PROOF OF (2). The argument is similar to that in the proof of theorem 1.
$\Leftarrow$ Choose a model $\mathcal{K}$ with $\mathcal{K} \models comp(P)$. We have $T_P^3(\mathcal{K}) \leq \mathcal{K}$:
(i) suppose r is assigned 1 by $T_P^3(\mathcal{K})$, then there exists a program clause $\theta \rightarrow r$ in P such that $\mathcal{K}$ assigns 1 to θ. Since $\mathcal{K} \models comp(P)$, in particular $\mathcal{K} \models r \leftrightarrow Def(r)$, and since $\theta \rightarrow Def(r)$, it follows that r is true on $\mathcal{K}$.
(ii) suppose r is assigned 0 by $T_P^3(\mathcal{K})$, then there exists a program clause $\theta \rightarrow r$ in P and for all such clauses, $\mathcal{K}$ assigns 0 to their bodies. It follows that $Def(r)$ is assigned

0 by $\mathcal{K}$, hence the same holds for r.
(iii) if r has value u in $T^3_P(\mathcal{K})$, this means neither (i) nor (ii) applies and there exists no program clause $\theta \to r$ in P with θ either 0 or 1. It follows that θ must have value u, hence r as well.
Note that we may have $T^3_P(\mathcal{K}) < \mathcal{K}$, for instance in case $P = \{q \to r\}$ and $\mathcal{K} \models$ *comp*(P), $\mathcal{K}$ makes r, q false, then $T^3_P(\mathcal{K})$ makes q undecided.
The set of models $\{\mathcal{B} \mid \mathcal{B} \leq \mathcal{K}\}$ is a dcpo, hence T^3_P has a least fixpoint $\mathcal{M} \subseteq \mathcal{K}$ here. Indeed, if **0** denotes the bottom element of the dcpo, then $\mathbf{0} \leq \mathcal{K}$ implies $T^3_P(\mathbf{0}) \leq T^3_P(\mathcal{K}) \leq \mathcal{K}$, whence it follows that the least fixpoint of T^3_P is a submodel of any $\mathcal{K}$ such that $\mathcal{K} \models$ *comp*(P). By hypothesis $\mathcal{M} \models F$. Since F is monotone and $\mathcal{M} \leq \mathcal{K}$, $\mathcal{K} \models F$, whence *comp*$(P) \models F$.
$\Rightarrow$ Since $\mathcal{M}$ is the least fixpoint of T^3_P, $\mathcal{M} \models$ *comp*(P), whence $\mathcal{M} \models F$. □

One step in the proof deserves special mention

Lemma 8. *For any model $\mathcal{K}$ with $\mathcal{K} \models$ comp(P) one has $T^3_P(\mathcal{K}) \leq \mathcal{K}$. In other words, a model of the completion is a pre-fixpoint of the consequence operator.*

A final remark regarding Lemma 4(3) in Chapter 7 of *Human Reasoning and Cognitive Science* is that it inadvertently stated that every model for the completion is a fixpoint. This doesn't affect the cognitive applications however, which are couched in terms of least fixpoints; and as we have seen entailment is determined by the least fixpoint.

References

1. Achourioti, T., Fugard, A., Stenning, K.: The empirical study of norms is just what we are missing. Frontiers in Cognitive Science 5, 1159, doi: 10.3389/fpsyg.2014.01159 (2014)
2. Baggio, G., van Lambalgen, M., Hagoort, P.: Logic as Marr's computational level: Four case studies. Topics in Cognitive Science (2014)
3. Batens, D.: A universal logic approach to adaptive logics. Logica Universalis 1, 221–242 (2007)
4. Bonnefon, J. F.: A theory of utility conditionals: paralogical reasoning from decision-theoretic leakage. Psychological Review 116, 888 – 907 (2009)
5. Bröder, A.: Decision making with the adaptive toolbox. Influence of environmental structure, intelligence and working memory load. Journal of Experimental Psychology: Learning, Memory and Cognition 29, 601-625 (2003)
6. Byrne, R. J: Suppressing valid inferences with conditionals. Cognition 3, 61 – 83 (1989)
7. Cummins, D. D.: Naïve theories and causal deduction. Memory and Cognition 23, 646 – 658 (1995)
8. Day, S. B., Gentner, D.: Nonintentional analogical inference in text comprehension. Memory and Cognition 35, 39 – 49 (2007)
9. Demolombe, R., Fernandez, A. M. O.: Intention recognition in the situation calculus and probability theory frameworks. In: Proceedings of Computational Logic in Multi-agent Systems (CLIMA) (2006)
10. Dougherty, M. R., Franco-Watkins, A. M., Thomas, R.: Psychological plausibility of the theory of probabilistic mental models and the fast and frugal heuristics. Psychological Review 115, 199 – 211 (2008)
11. Einstein, A.: Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. Annalen der Physik 4, 132-147 (1905)
12. Gergely, G., Bekkering, H., Király, I.: Rational imitation in preverbal infants. Nature 415, 755 (2002)
13. Gigerenzer, G., Todd, P. M., the ABC research group: Simple heuristics that make us smart. Oxford University Press, New York (1999)
14. Hertwig, R., Benz, B., Krauss, S.: The conjunction fallacy and the many meanings of and. Cognition 108, 740 – 753 (2008)
15. Hölldobler, S., Kencana Ramli, C. D. P.: Logic Programs under Three-Valued Lukasiewicz's Semantics. In: P.M. Hill and D.S. Warren (eds.), Logic Programming, LNCS series, vol. 5649, pp. 464 – 478, Springer-Verlag Berlin Heidelberg (2009)
16. Holte, R. C.: Very simple rules perform well on mostly used data sets. Machine Learning 11, 63 – 91 (1993)
17. Henle, M.: On the relation between logic and thinking. Psychological Review 69, 366 – 378 (1962)
18. Hertwig, R., Benz, B., Krauss, S.: The conjunction fallacy and the many meanings of and. Cognition 108, 740 – 753 (2008)
19. Király, I., Csibra, G., Gergely, G.: Beyond rational imitation: Learning arbitrary means actions from communicative demonstrations. Journal of Experimental Child Psychology 116, 471 – 486 (2013)

20. Kowalski, R.: The Early Years of Logic Programming. In: Commun. ACM, vol. 31, pp. 38 – 43 (1988)
21. Kowalski, R.: Computational Logic and Human Thinking: How to be Artificially Intelligent. Cambridge University Press, New York Cambridge University Press, New York (2011)
22. Marevski, J., Schooler, L. J.: Cognitive niches: an ecological model of strategy selection. Psychological Review 118, 393 – 437 (2001)
23. Martignon, L., Hoffrage, U.: Fast, frugal and fit: Simple heuristics for pair comparison. Theory and Decision 52, 29 – 71 (2002)
24. Martignon, L., Stenning, K., van Lambalgen, M.: Qualitative models of reasoning and judgment: theoretical analysis and experimental exploration (in prep)
25. McCarthy, J., Hayes, P.: Some philosophical problems from the standpoint of Artificial Intelligence. In: Machine Intelligence, pp. 463-502, Edinburgh University Press (1969)
26. Meltzoff, A. N.: The 'like me' framework for recognizing and becoming an intentional agent. Acta Psychologica 124, 106 – 128 (2007)
27. Payne, J. W., Bettman, J. R., Johnson, E. J.: The adaptive decision maker. Cambridge University Presss, Cambridge (1993)
28. Pearl, J.: Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufman Publishers, San Francisco (2014)
29. Politzer, G.: Premise interpretation in conditional reasoning. In: Hardmann, D., Macchi, L. (eds.), Thinking: Psychological Perspectives on Reasoning, Judgment, and Decision Making, pp. 79 – 93 (2003)
30. Rieskamp, J., Otto, P. E.: SSL: A theory of how people learn to select strategies. Journal of Experimental Psychology: General 135, 207 – 236 (2006)
31. Sadri, F.: Logic-based approaches to intention recognition. In: Handbook of Research on Ambient Intelligence: Trends and Perspectives, pp. 346 – 375, Citeseer (2010)
32. Stenning, K., van Lambalgen, M.: Human Reasoning and Cognitive Sciences. MIT University Press, Cambridge, MA (2008)
33. Stenning, K., Varga, A.: Many logics for the many things that people do in reasoning. In: International Handbook of Thinking and Reasoning, Ball, L., Thompson, V. (eds.), Psychology Press, London (under review)
34. van Lambalgen, M., Hamm, F.: The Proper Treatment of Events. Blackwell Publishing, Malden (2005)
35. Varga, A.: A Formal Model of Infants' Acquisition of Practical Knowledge from Observation. PhD thesis, Central European University, Budapest (2013)
36. Varga, A., Gazzo Castañeda, E.: Yes, you may Affirm the Consequent – In a conversation setting (in prep)