

From Logic Programming to Human Reasoning: How to be Artificially Human

Emmanuelle-Anna Dietz Saldanha
Knowledge Representation and Reasoning Group
Technische Universität Dresden, 01062 Dresden, Germany
emmanuelle.dietz@tu-dresden.de

Zusammenfassung

Das Thema der Dissertation befindet sich an der Schnittstelle der Kognitionswissenschaften und der Computational Logic. Wir besprechen verwandte Arbeiten in beiden Gebieten und geben eine Übersicht der Stand der Wissenschaft. Das Ziel dieser Arbeit ist es eine Methodologie zu entwickeln die es uns erlaubt, Episoden menschlichen Schließens im Bezug auf Konditionale adäquat zu modellieren. Dafür haben wir Ergebnisse entsprechender psychologischer Experimente untersucht und mithilfe der Weak Completion Semantics, einer Herangehensweise in der Logik Programmierung, formalisiert. Außerdem stellen wir ein allgemeines Evaluationssystem für Konditionale vor.

1 Einführung

Ergebnisse psychologischer Experimente haben gezeigt, dass Menschen Annahmen machen die nicht notwendigerweise gültig sein müssen, dass sie sich von ihrem Hintergrundwissen beeinflussen lassen und dass sie nicht-monotone Schlüsse ziehen. Das sind Indizien dafür, dass klassische Logik nicht geeignet ist für das Modellieren von menschlichem Schließen. Betrachten wir folgendes Beispiel aus einem psychologischen Experiment mit Syllogismen, dass von Evans, Barston, and Pollard [1983] ausgeführt wurde. Teilnehmer sollten entscheiden, ob der folgende Syllogismus gültig ist:

- PRÄMISSE 1 *No millionaires are hard workers.*
(Keine Millionäre sind hartarbeitend.)
- PRÄMISSE 2 *Some rich people are hard workers.*
(Manche reiche Menschen sind hartarbeitend.)
- SCHLUSS *Therefore, some millionaires are not rich people.*
(Also sind manche Millionäre keine reichen Menschen.)

Ist dieser Syllogismus gültig? Gültig zu sein heißt, dass wenn die ersten zwei Prämissen als wahr angenommen werden, der Schluss notwendigerweise daraus folgen muss. Die Mehrheit der Befragten antworteten, dass dieser Syllogismus nicht gültig sei. Die Befragten sollten auch den folgenden Syllogismus bewerten:

- PRÄMISSE 1 *No addictive things are inexpensive.*
(Keine süchtig machenden Dinge sind preiswert.)
- PRÄMISSE 2 *Some cigarettes are inexpensive.*
(Manche Zigaretten sind preiswert.)
- SCHLUSS *Therefore, some addictive things are not cigarettes.*
(Also sind manche süchtig machenden Dinge keine Zigaretten.)

In diesem Fall, kam die Mehrheit der Befragten zu dem Schluss, dass der Syllogismus gültig sei. Nun verhält es sich wie folgt: wenn man beide Syllogismen genauer betrachtet, dann sieht man, dass ihre Strukturen identisch sind, das bedeutet, dass entweder beide gültig sein müssen oder es beide nicht sind. Allerdings stimmt dies nicht mit der Mehrheit der Antworten der Befragten überein.

Evans, Barston und Pollard wollten mit diesem Experiment zeigen, dass Menschen vom sogenannten Überzeugungsbias beeinflusst werden. Der Überzeugungsbias beschreibt das Phänomen, welches auftritt, wenn wir glauben etwas auf Basis unseres logischen Denkens zu beurteilen, allerdings unbewusst von unserer persönlichen Meinung, beziehungsweise unserem Hintergrundwissen beeinflusst werden.

In dem Fall der zwei oben besprochenen Beispiele sind beide Syllogismen in der klassischen Logik nicht gültig. Man kann dies am Besten mithilfe eines Venn-Diagramms zeigen. Betrachten wir das Venn-Diagramm in Abbildung 1. Die erste Prämisse *Keine süchtig machenden Dinge sind preiswert* lässt uns schlussfolgern, dass die

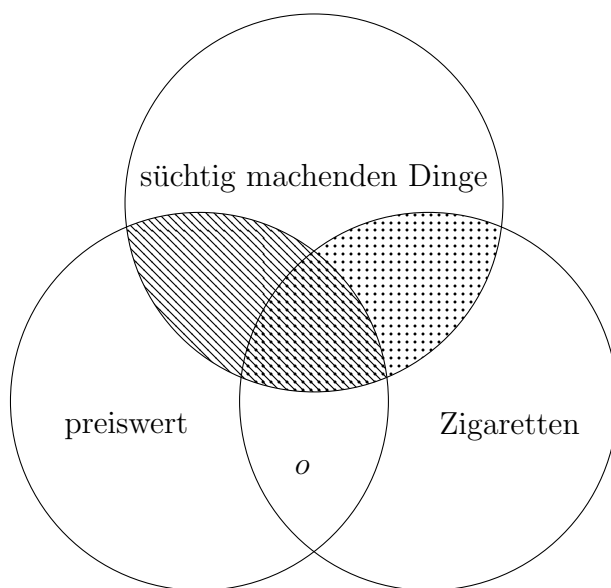


Abbildung 1: Venn-Diagramm, das zeigt, dass beide Syllogismen nicht gültig sind.

Überschneidung zwischen süchtig machende Dinge und preiswert leer ist, das heißt, es gibt nichts, das sowohl süchtigmachend als auch preiswert ist, angegeben mit den schraffierten Linien. Anhand der zweiten Prämisse wissen wir, dass es etwas gibt, das preiswert ist und eine Zigarette, zum Beispiel ein Objekt o . Das ist angegeben durch das o in der Überschneidung von preiswert und Zigaretten. Wir versuchen den Schluss *Also sind manche süchtig machenden Dinge keine Zigaretten* zu ziehen, indem wir überprüfen, ob es mindestens ein Objekt gibt, das süchtig machend ist, aber keine Zigarette. Das ist tatsächlich nicht der Fall, denn alle süchtig machenden Dinge könnten ja einfach die Zigaretten sein, angegeben durch die Punkte in der Überschneidung zwischen süchtigmachenden Dingen und Zigaretten. Somit ist der Syllogismus nicht gültig. Analog kann auch der andere Syllogismus überprüft werden. Betrachten wir noch einmal folgende Prämisse aus dem vorher besprochenen Syllogismus:

No addictive things are inexpensive.
(Keine süchtig machenden Dinge sind preiswert.)

Daraus kann man klassisch logisch nicht schließen, dass es überhaupt etwas gibt, das süchtig machend ist. Allerdings ist es so, dass Menschen bei der Allquantifizierung

von Dingen, also zum Beispiel bei Aussagen wie *für alle süchtig machenden Dinge* beziehungsweise *für keine süchtig machenden Dinge*, implizit annehmen, dass es süchtig machende Dinge gibt.

Statt anzunehmen, dass Menschen gar nicht logisch schließen, wird in dieser Arbeit davon ausgegangen, dass Menschen nicht *klassisch* logisch schließen. Dafür wird ein nicht klassisch logischer Ansatz, die sogenannte Weak Completion Semantics, ursprünglich vorgestellt in [Hölldobler and Kencana Ramli, 2009a,b], herangezogen. Die Weak Completion Semantics ist eine Herangehensweise aus der Logikprogrammierung bei der das kleinste Modell der schwachen Vervollständigung eines Logikprogramms unter der dreiwertigen Lukasiewicz [1920] Logik betrachtet wird. Bei der dreiwertigen Lukasiewicz Logik gibt es zusätzlich zu den zwei Wahrheitswerten aus der klassischen Logik, *wahr* und *falsch*, noch den dritten Wahrheitswert, *unbekannt*.

2 Weak Completion Semantics

Da die Weak Completion Semantics eine relativ neue und noch nicht umfassend untersuchte Herangehensweise ist, motivieren wir im ersten Teil der Arbeit weshalb dieser Ansatz interessant für das Modellieren von menschlichem Schließen ist. Ausserdem hat die Weak Completion Semantics formale Eigenschaften die vorteilhaft sind: Jedes Programm hat ein kleinstes Modell und dieses kann mithilfe eines semantischen Fixpunktoperator berechnet werden.

Zusätzlich zur Weak Completion Semantics setzen wir auch Integritätsbedingungen und Abduktion ein. Abduktion ist eine nicht klassisch logische Form des Denkens die häufig bei menschlichen Schließen beobachtet wird. Betrachten wir das folgende Beispiel:

Wenn es regnet, dann ist das Gras nass.
Wenn der Rasensprenger an ist, dann ist das Gras nass.

Nehmen wir an, dass *es regnet*. Es ist klassisch logisch korrekt, dass wir daraus schließen können, dass *das Gras nass* sei. Das nennt man Deduktion.

Wenn wir jetzt aber beobachten sollten, dass *das Gras nass* ist, dann können wir dies damit erklären, dass *es regnet*. Das nennt man Abduktion. Obwohl wir im Alltag häufig Abduktion verwenden, ist das kein klassisch logisch korrektes Schließen, denn

es könnte auch andere Erklärungen zu unserer Beobachtung geben, zum Beispiel weil *der Rasensprenger an ist*.

Nachdem wir im ersten Teil der Arbeit eine Einführung der Weak Completion Semantics und den zusätzlichen Techniken liefern, untersuchen wir im Anschluss schon etablierte Herangehensweisen der Logikprogrammierung. Dabei betrachten wir unter anderem die Well-founded Semantics [Van Gelder, Ross, and Schlipf, 1991] und die Stable Model Semantics [Gelfond and Lifschitz, 1988] und zeigen die formale Korrespondenz zur Weak Completion Semantics. Zusätzlich geben wir einen Überblick zu verwandten Semantiken und zeigen die Zusammenhänge zu bekannten zweiwertigen Semantiken, sowie die (Well)-supported Model Semantics und Clark's Completion Semantics.

Es gibt zwei Hauptunterschiede zwischen der Weak Completion Semantics und der Well-founded Semantics: Der erste Unterschied liegt im Verfahren mit undefinierten Atomen in Logikprogrammen, bei der die Well-founded Semantics und die Stable Model Semantics die Annahme zur Weltabgeschlossenheit annehmen, und die Weak Completion Semantics stattdessen diese Annahme nicht macht. Der zweite Unterschied bezieht sich auf dichte Programme: Unter bestimmten Umständen sind Atome die in einem positiven Zyklus beteiligt sind, falsch unter der Well-founded Semantics, wobei sie unbekannt unter der Weak Completion Semantics bleiben. Da die Weak Completion Semantics und die Well-founded Semantics unterschiedlich mit positiven Zyklen in Logikprogrammen umgehen, haben wir eine psychologische Studie über zyklische Konditionale ausgeführt. Es scheint, dass Teilnehmer positive Zyklen in Konditionalen der Länge 1, als Fakten verstanden haben. Bei Zyklen in Konditionalen der Länge 2 oder 3, haben sie diese als tatsächliche Konditionale verstanden. Diese Voruntersuchung zeigt, dass das Verständnis der Teilnehmer bezüglich der positiven Zyklen ähnlich ist, wie die Weak Completion Semantics mit Zyklen umgeht.

Zum Schluss des ersten Teils stellen wir eine Erweiterung der Weak Completion Semantics vor und führen kontextabhängige Logikprogramme ein. In dieser Erweiterung gibt es einen zusätzlichen Operator, den *Kontext*, der eine Abbildung von einer Dreiwertigkeit auf eine Zweiwertigkeit ermöglicht. Ausserdem können wir mit diesem Operator schwache Negation unter der Weak Completion Semantics ausdrücken. Wir überprüfen ob die formalen Ergebnisse der Weak Completion Semantics für Logikprogramme auf kontextabhängige Logikprogramme erweitert werden können. Dabei zeigen wir, dass der semantische Fixpunktoperator für kontextabhängige Programme nicht mehr monoton ist. Außerdem hat der semantische Fixpunktoperator nicht un-

bedingt einen kleinsten Fixpunkt für diese Klasse von Programmen. Nichtsdestotrotz, können wir garantieren, dass ein kleinster Fixpunkt existiert, wenn wir die Klasse der Programme auf azyklische kontextabhängige Programme einschränken. Danach stellen wir eine Herangehensweise der kontextabhängigen Abduktion vor, die es uns erlaubt, Präferenzen unter den Erklärungen auszudrücken. Diese Herangehensweise erlaubt es auch andere Beziehungen zwischen Beobachtungen auszudrücken. Wir spezifizieren, wenn eine Beobachtung ein kontextabhängiger (anfechtbarer) Nebeneffekt einer anderen Beobachtung ist, oder ob beide zusammen gestützte kontextabhängige Konsequenzen voneinander sind.

3 Experimente zum Menschlichen Schließen

Im zweiten Teil der Arbeit untersuchen wir erst Byrne's [1989] Suppression Task und Wason's [1968] Selection Task. Wir prüfen die Suppression Task bezüglich der Well-founded Semantics und zeigen, dass die angegebenen Logikprogramme verändert werden müssen, um die Ergebnisse adäquat unter der Well-founded Semantics zu modellieren. Danach betrachten wir ein Experiment aus [Ragni and Knauff, 2013] zum räumlichen Schließen und modellieren es anhand von den Ideen der Preferred Model Theory [Ragni and Knauff, 2013]. Dafür entwickeln wir eine Rapresentation der first free fit technique in Logikprogrammen: Die Idee ist, dass wenn ein neues Objekt hinzukommt, es entweder direkt neben dem schon existierenden Objekt hinzugefugt wird, gegeben dass dort Platz ist, oder falls nicht, wird das neue Objekt zum nachsten freien Platz hinzugefugt. Psychologische Experimente haben gezeigt, dass diese Technik kognitiv adquat ist. Unsere Ergebnisse zum raumlichen Schließen stimmen mit den Ergebnissen der psychologischen Experimente uberein.

Danach entwickeln wir funf Prinzipien zur Rapresentation von quantifizierten Aussagen motiviert durch die Techniken aus der Logikprogrammierung und den Erkenntnissen der Kognitionswissenschaften. Wir stellen danach eine Rapresentation der vier moglichen quantifizierten Aussagen als Logikprogramme vor. Im Anschluss berechnen wir fur die 64 moglichen syllogistischen Pramissen unsere Vorhersagen und vergleichen sie mit den Ergebnissen psychologischer Experimente von Khemlani and Johnson-Laird [2012]. Das Ergebniss dieser Vorhersagen sticht hier hervor da wir eine ubereinstimmung von 85% mit den Antworten der Teilnehmer erzielen. Das ist mehr, als das was die 12 aktuellen kognitionswissenschaftlichen Theorien voraussagen konnen. Zum Schluss betrachten wir das in der Einfuhrung betrachtete psychologi-

sche Experiment von Evans, Barston, and Pollard [1983], dass den Überzeugungsbias in syllogistischen Aussagen testet. Dafür nehmen wir die vorherigen entwickelten Prinzipien als Ausgangspunkt. Wie stellen fest, dass der Überzeugungsbias entweder in der Repräsentationsphase oder in der Schlussfolgerungsphase entstehen kann. Dazu entwickeln wir zwei zusätzliche Prinzipien und modellieren den Überzeugungsbias mithilfe von Abnormalitätspredikaten und Abduktion.

Nach bestem Wissen, kennen wir keine Theorie die gezeigt hat so viele unterschiedliche Episoden menschlichen Schließens modellieren zu können.

4 Konditionale

Der letzte Teil der Arbeit präsentiert ein abstraktes Evaluationssystem für Konditionale. Wir führen einen Revisionsoperator ein zusammen mit Abduktion und zeigen anhand von bekannten Beispielen aus der Literatur, dass wir praktisch alle Ergebnisse modellieren können, d.h. gegeben ein Konditional, können wir es entweder als *falsch*, *wahr*, oder *unbekannt* auswerten. Das Ergebnis hängt hauptsächlich davon ab, ob erst Revision oder Abduktion angewendet wird, und in welcher Reihenfolge die Bedingungen im Konditional behandelt werden. Außerdem diskutieren wir in diesem Kontext das Thema der Relevanz im Zusammenhang mit Konditionalen. Wir unterscheiden zwischen einer schwachen und einer starken Relevanz und schlagen allerdings dann eine dritte, nämlich eine semantische Definition von Relevanz vor, da die ersten zwei Definitionen uns in diesem Kontext nicht adäquat erscheinen. Wir vermuten, dass Menschen Abduktion der Revision vorziehen, was auch mit den Annahmen der Literatur in der Kognitionswissenschaft übereinstimmt. Allerdings fehlen uns für eine sorgfältige Auswertung experimentelle Daten. Zum Schluss zeigen wir, dass unser System allgemeiner ist als ein anderes System [Schulz, 2014] dass kürzlich in der Literatur vorgestellt wurde.

5 Fazit und Ausblick

Zusammenfassend, haben wir einen möglichen Weg aufgezeigt, wie man Resultate aus den Kognitionswissenschaften und Techniken aus der Computational Logic miteinander

der kombinieren kann. Die Ergebnisse dieser Dissertation sollen nicht als endgültige Lösung betrachtet werden. Vielmehr können die neu gewonnen Erkenntnisse einen Ausgangspunkt bieten und als Orientierungshilfe für neue Fragen zwischen beiden Gebieten, der Kognitionswissenschaften und der Computational Logic, dienen.

Literatur

- R. M. Byrne. Suppressing valid inferences with conditionals. *Cognition*, 31:61–83, 1989.
- J. S. Evans, J. L. Barston, and P. Pollard. On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, 11(3):295–306, 1983.
- M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In R. Kowalski and K. A. Bowen, editors, *Proceedings of the International Logic Programming Conference and Symposium, (ICLP/SLP 1988)*, pages 1070–1080, Cambridge, MA, 1988. MIT Press.
- S. Hölldobler and C. D. Kencana Ramli. Logic programs under three-valued Lukasiewicz semantics. In P. M. Hill and D. S. Warren, editors, *25th International Conference on Logic Programming, (ICLP 2009)*, volume 5649 of *Lecture Notes in Computer Science*, pages 464–478, Heidelberg, 2009a. Springer.
- S. Hölldobler and C. D. Kencana Ramli. Logics and networks for human reasoning. In C. Alippi, M. M. Polycarpou, C. G. Panayiotou, and G. Ellinas, editors, *International Conference on Artificial Neural Networks, (ICANN 2009), Part II*, volume 5769 of *Lecture Notes in Computer Science*, pages 85–94, Heidelberg, 2009b. Springer.
- S. Khemlani and P. N. Johnson-Laird. Theories of the syllogism: A meta-analysis. *Psychological Bulletin*, 138(3):427–457, 2012.
- J. Łukasiewicz. O logice trójwartościowej. *Ruch Filozoficzny*, 5:169–171, 1920. English translation: On three-valued logic. In: Łukasiewicz J. and Borkowski L. (ed.). (1990). *Selected Works*, Amsterdam: North Holland, pages 87–88.
- M. Ragni and M. Knauff. A theory and a computational model of spatial reasoning with preferred mental models. 120(3):561 – 588, 2013.

- K. Schulz. Minimal models vs. logic programming: the case of counterfactual conditionals. *Journal of Applied Non-Classical Logics*, 24(1-2):153–168, 2014.
- A. Van Gelder, K. A. Ross, and J. S. Schlipf. The well-founded semantics for general logic programs. *Journal of the ACM*, 38(3):619–649, 1991.
- P. C. Wason. Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20(3):273–281, 1968.