Modeling Arguments about the Curry Paradox using Argumentation Theory

Dominic Deckert
Diplom course in Computer Sciences
Technical University Dresden
dominic.deckert@mailbox.tu-dresden.de

1 Introduction

Computers can reason. This thesis statement is be one of the central ideas of knowledge representation and resasoning. If given knowledge about an area in a formalized and comprehensive form, computers are able to arrive at inferences that (at least to some extent) mirror the conclusions a human would come to. This is easy to achieve in areas that are already highly abstracted and formalized, but more ambiguous cases pose a larger challenge – in both modeling and inference.

Argumentation Theory is a formalism that offers a generalized approach to present topics and different perspectives on them. This is especially true for debates, where these different points of view, arguing for their respective sides and undermining the validity of the other perspective, can represent topics that are difficult to navigate. An argumentation framework judges opposing (and thus, mutually inconsistent) perspectives to ascertain what facts and conclusions can be admitted. And like many other KRR approaches, most argumentation frameworks rely on the syntax of logic to represent their information.

But then again, what is logic? This question is asked (and answered) in many different fields – chief among them philosophy, mathematics and computer sciences. And in these different fields researchers might arrive at different answers. While computer sciences and mathematics mostly study highly formalized, syntactically and semantically defined and very abstract logics (e.g. FOL, modal logic and Description logics) other fields might study logical statements in natural language where such explicit definitions are harder to come by. By its very nature, natural language blurs the line between logical statements and statements about them.

When studying a logic, one of its most important properties is consistency (or in a more general framing, nontriviality[1]). Nothing can be gained from reasoning in a trivial logic where all statements can be derived in some way. To avoid this, philosophical logic studies logic paradoxes and provides arguments to refute them. The Curry paradox is one

[1] In *paraconsistent logics*, a field of philosophical logics, there is a difference between a logic that can derive falsity (and is thus inconsistent), and one that can derive any statement (and is thus trivial).

such well-researched paradox in the logic of natural language. In the field of philosophy, different approaches (and counter-arguments) have been presented to refute the Curry paradox.

In this paper, we want on one hand to provide a different perspective by looking at how lay people deal with the Curry paradox. In 2018, Cramer and Guillaume conducted a survey on counter-arguments to the Curry paradox, letting lay people formulate and discuss their responses to the paradox. We would like to provide an initial overview of the survey by chosing and presenting a few insightful responses. On the other hand, this paper is intended to be proof of concept for the ASPIC-END argumentation framework. Based on the well-researched ASPIC$^+$ framework, this formalism utilizes the additional notion of explanations to represent disputes between different arguments and find positions that can reasonably be justified. With ASPIC-END, we can provide more formalized versions of the chosen survey responses.

In section 2.1, we will present an overview of the Curry paradox. Section 2.2 will provide an introduction to ASPIC-END, while in section 2.2.3 will introduce new *Meta-Language References* to ASPIC-END. In section 3 we will present the survey and formalize the chosen responses in ASPIC-END.

2 Preliminaries

In this section, we will present the important notions for this paper. First, we will detail what the Curry paradox is, where it can arise and a few common responses to it. The later part of this section will deal with the modeling of our argument. We have chosen the ASPIC-END argumentation framework for this task. ASPIC-END is an extended form of the well-known ASPIC$^+$ argumentation framework ([MP14]), enhanced with explanatory relations ([ŠS13]) and some new argument-related features (such as reasoning by cases or contradictions). Lastly, we will propose extensions to the ASPIC-END formalism to incorporate meta-language statements into the language and allow reasoning about the structure of arguments (in arguments).

2.1 The Curry paradox

"Logic", as a concept, can refer to different things. For the mathematician, this is mostly the study of propositional and First Order logic (and their logic families), where logical statements follow an explicit syntax and can (usually) be evaluated to truth or falsehood following an explicit semantics.

In philosophical logic, the studied logics encompass a much wider area of possibility. Anything that assigns statements with truth values in an internally consistent fashion can be considered a logic theory ([SW20]). This includes analysis of statements in quasi-natural language (with some syntactical structure).[2] Capturing the notions of logic – i.e. validity, formal consequences and others – in natural language presents the opportunity to make the logic and its inferences widely understandable. This way, human reasoning

[2] This idea originated from a paper by Polish logician Alfred Tarski, see [T$^+$56]

(and human thinking in general) can be captured and discussed easily.
One of the focuses in the study of such logic theories is the way in which paradoxical statements can be constructed and in which ways they can be refuted in certain logical theories. One such paradox is the Curry's paradox (see [SB18]).
The Curry paradox centers around a statement S of the form "*If this sentence S is true, then C follows*" (this can also be written as $S = S \to C$) where C can be any statement, however absurd. Under standard logical semantics, one can show that this statement must be true (and that thus any C must be true). The proof can usually be made in two steps: first show that the implication in S must hold, and then conclude C. To prove an implication, we can use the principle of *implication introduction.* If, by assuming the antecedent S of an implication, we can deduce the conclusion C, we have then shown that the whole implication holds. Assuming that the antecedent (i.e. S) holds, we can conclude by *modus ponens* on S and $S \to C$ that C must also hold. With this, we can conclude that the implication (i.e. S) holds and that thus (again by modus ponens) C must hold as well.

For implication introduction, we assume S:	S
By definition of S, this is equivalent to:	$S \to C$
By modus ponens on the above two statements:	C
Thus, implication introduction is successful, and we conclude:	$S \to C$
By definition of S, this is equivalent to:	S
Again by modus ponens on the above two statements:	C

Table 1: Outline of the standard Curry argument

As a note: traditionally, the Curry sentence concludes that "the moon is made of cheese," but one can just as easily construct a Curry sentence for any conclusion. Such a logic is called *trivial* and is useless in the sense that we cannot formally distinguish between valid conclusions and inane derivations. While in classical logic this property holds even for any inconsistent logic, philosophers have proposed *paraconsistent* logics where inconsistencies do not trivialize a logic. This makes triviality a strictly more damaging property, and exacerbates the problem the Curry paradox poses.
More generally speaking, the Curry paradox is a problem of meta-statements: in some natural language scenarios, one may wish to speak of some property of statements themselves.[3] This arises from the fact that in natural language, it is difficult to clearly delineate between the *object language*, used to describe the matter at hand, and the *meta-language*, used for reasoning about statements of the object language (see [Hod18] for further elaborations on this distinction). Considering that both are expressed using natural language,[4] one might be interested in cases where a statement reasons about its own truth value, which is also the basis for the well-known *Liar's paradox* (which is centered around the statement "This statement is false").

[3] i.e. "Any statement of the structure 'If A, then A' is trivially true"

[4] Which means that natural language is powerful enough to be its own meta-language

2.1.1 Responses to the Curry paradox

The existence of a Curry sentence in a logic is obviously problematic, since a logic that can conclude any statement is trivial (see [SB18]). To reconcile our hopes for useful natural language logics with the existence of a Curry sentence, philosophers have tried to formulate responses to the Curry paradox. These responses fall in several categories (for a more in-depth presentation of the different categories see [SB18]), the most fundamental of which either strengthen the divide between object and meta language or attack the reasoning principles necessary to derive the Curry sentence's truth value. Responses in the first category reject Curry sentences outright for the fact that they refer to the meta-property of a sentence (its truth value). And indeed, some logics that are powerful enough to express complex concepts (such as arithmetics) run afoul of e.g. Tarski's *Undefinability Theorem*, which states that in such a logic *truth* of a statement cannot be defined.

The other category of approaches state that one of the reasoning steps (e.g. the modus ponens) cannot be applied to the Curry sentence in the way that is necessary to derive C. As an example, some philosopher's reject the "structural contraction" happening as part of deriving the Curry sentence: during the proof of the Curry sentence, the premise of S had to be taken "twice" (once to prove the Curry sentence and then to use modus ponens again). These two uses of the premise are contracted in the course of proving the Curry sentence, and specifically, applying the modus ponens (see for example [BM13]). Another interesting approach that settles close to the first category is a *Truth-value gap* explanation[5]: Since S is a self-referencing sentence that refers to its own truth value in such a way that assuming any truth value for it leads to a problematic result, it could be argued that S should be assigned neither truth value, and instead be left to the "gap" between traditional truth values, having neither value.

2.2 Modeling in ASPIC-END

In the last section, we have presented a rough overview of the Curry paradox. This included a foray into the different philosophical responses to the paradox. As mentioned in the introduction, we would like to present some of these responses – as well as the Curry paradox itself – in a unified manner. To this end we have chosen to use a Dung-style *argumentation framework*, specifically the ASPIC-END formalism, an extension on the well-researched ASPIC$^+$ framework. Dung-style argumentation frameworks in general can be used to express arguments and how they might rebut or otherwise attack one another. Using this network of attacking and attacked arguments, one can use argumentation frameworks to find justifiable positions in a given discourse.

While ASPIC$^+$ judges the acceptability of arguments on how well they are defended from counter-arguments, the ASPIC-END framework makes use of *explanatory argumentation frameworks* ([ŠS13]). This formalism expands the general argumentation framework approach by judging acceptability on how well arguments can explain certain facts, so-called *explananda*.

[5] See [SW20] on truth-value gaps in general

The formal definitions are made in accordance with the introductory paper for ASPIC-END ([CD20]) and more extensive definitions can be found there.

2.2.1 Argumentation Frameworks

The notion of an argumentation framework was first introduced by Dung in [Dun95]. While his abstract argumentation frameworks do not represent the inner structure of an argument, they did however express how some arguments might "attack" and thus weaken belief in other arguments. He also provided ways to define semantics that accept arguments (or sets of arguments, so-called *extensions*) on the grounds of whether they are attacked, and whether these attacks can be defended against.
Multiple formalisms have been proposed for constructing arguments and their relations with regards to their inner structure. One of the most well-known (and well-researched) is the ASPIC$^+$ framework, on which the ASPIC-END formalism is based (see [MP14] for an introduction). In section 2.2.2, we will define ASPIC-END.
Additionally, ASPIC-END induces a special *explanatory argumentation framework* (EAF), which utilizes the notion of *explanations* (the idea for EAFs was introduced in [ŠS13]). Arguments (or rather *extensions*) are judged on how well they explain a given set of *explananda* (i.e. observations that need to be explained) as well.

Definition 1. *An* explanatory argumentation framework *is a 4-tuple* $(\mathcal{A}, \mathcal{X}, \longrightarrow, \dashrightarrow)$ *where*

- $\mathcal{A}$ *is a set of arguments*
- $\longrightarrow \subseteq \mathcal{A} \times \mathcal{A}$ *is the* attack *relation*
- $\mathcal{X}$ *is the set of explananda and*
- $\dashrightarrow \subset \mathcal{A} \times (\mathcal{A} \cup \mathcal{X})$ *is the* explanation *relation*

It is important to note that $(\mathcal{A}, \longrightarrow)$ is a Dung-style abstract argumentation framework. In this formalism (and in EAFs as well) we will focus mostly on extensions that are particularly justified. These extensions are called admissible:

Definition 2. *An extension* $\mathcal{T} \subseteq \mathcal{A}$ *is called* admissible *if*

- *every argument* $o \in \mathcal{T}$ *is* defended*, i.e. for every* n *with* $n \longrightarrow o$ *there is a* $m \in \mathcal{T}$ *such that* $m \longrightarrow n$*, and*
- $\mathcal{T}$ *is* conflict-free*, i.e. there are no* $m, n \in \mathcal{T}$ *such that* $m \longrightarrow n$

The admissible extensions are the basis for any semantics in abstract argumentation frameworks and EAFs. We only consider positions that cannot be refuted by other arguments and are internally consistent. In addition, we expand the general Dung-style approach by judging the acceptability of positions by how well they provide explanations for the explananda of the EAF.

Definition 3. *An* explanation *for an explanandum E and an admissible extension $\mathcal{T}$ is a set $X[E] \subset \mathcal{T}$ such that there is an unique $A \in X[E]$ such that $A \dashrightarrow E$ and for all other arguments $A' \in X[E] \setminus \{A\}$ there is a path from A' to A in $\dashrightarrow$.*
The set of all explananda explained by an extension $\mathcal{T}$ can be denoted as $\mathcal{E}_{\mathcal{T}}$.

An explanation can provide a step-wise explanation from (ideally) simple arguments to a final argument that then explains the explanadum directly. To differentiate the quality of an extension, we distinguish them by how many explananda they explain and how extensive they are:

Definition 4. *Given two admissible extension $\mathcal{T}, \mathcal{S}$, we say that:*

- *the extension $\mathcal{T}$ is* explanatory more powerful *than $\mathcal{S}$ ($\mathcal{T} >_p \mathcal{S}$) if $\mathcal{E}_{\mathcal{T}} \supsetneq \mathcal{E}_{\mathcal{S}}$*
- *the extension $\mathcal{T}$ is* explanatory deeper *than $\mathcal{S}$ ($\mathcal{T} >_d \mathcal{S}$) if for every explanation X' offered by $\mathcal{S}$, there is an explanation X offered by $\mathcal{T}$ such that $X \supseteq X'$ and there is at least one such pair with $X \supsetneq X'$*

From these notion, we can construct semantics that select only the most powerful and the deepest extensions:

Definition 5. *An admissible extension S is*

- satisfactory *if there is no admissible S' such that $S' >_p S$*
- insightful *if there is no satisfactory S' such that $S' >_d S$*

Furthermore, we can single out the extensions explaining the most explananda (called the *argumentative core extensions*), i.e. the set-maximal satisfactory extensions, and the extensions containing only the absolutely necessary arguments (called the *explanatory core extensions*), i.e. the set-minimal satisfactory extensions. These extensions provide useful standards to refer to for accepting arguments.

2.2.2 ASPIC-END

While the abstract and explanatory argumentation frameworks provide ways to treat relations between abstracted arguments, they do not actually model the contents of scientific discourse. This falls to ASPIC-END.[6] Central to ASPIC-END are the *argumentation theories* that can be defined in it. An argumentation theory represents the knowledge underlying a discourse in the form of two types of inference rules: *intuitively strict* rules that try to capture any "unbendable laws" that might exist in the knowledge base. The rules of logic, mathematics and physics are good examples: if we for example believe that something is true, then it must be the case that its negation is false. *Defeasible* rules on the other hand provide inferences that should hold under normal circumstances: a simple example would be the belief that "if something is a dog, then it has four legs."

[6] We heavily base the theoretical foundations of ASPIC-END on the foundational paper [CD20] by Cramer and Dauphin. All definitions can be read up on there.

Definition 6. *An* argumentation theory *is a tuple* $(\mathcal{L}, \mathcal{R}, n, <)$ *where:*

- $\mathcal{L}$ *is a logical language of free variables* $\mathcal{L}_v$*, function and relation symbols and closed under the logical connectives* $\vee$*,* $\neg$ *and* $\exists$*. Furthermore,* $\mathcal{L}$ *contains three special unary predicates* $\mathsf{Assumable}_\neg, \mathsf{Assumable}_\vee, \mathsf{Assumable}_\supset$ *and the nullary predicate* $\bot$ *(which represents triviality).*
- $\mathcal{R} = \mathcal{R}_{is} \cup \mathcal{R}_{def}$ *is the set of rules, consisting of a set of intuitively strict rules* $\mathcal{R}_{is}$ *of the form* $\phi_1, ..., \phi_n \rightsquigarrow \phi$ *and a set of defeasible rules* $\mathcal{R}_{def}$ *of the form* $\phi_1, ..., \phi_n \Rightarrow \phi$ *(in both cases,* $n \geq 0$ *and all* $\phi_i, \phi \in \mathcal{L}$*).*
  To represent triviality, we require that the rule set $R_\bot = \{\bot \rightsquigarrow \phi \mid \phi \in \mathcal{L}\} \subset \mathcal{R}_{is}$
- $n : \mathcal{R} \to \mathcal{L}$ *is a (partial) naming function (*n *is undefined for all* $r_\phi \in R_\bot$*)*
- $<$ *is an ordering relation over* $\mathcal{R}_{def}$ *that expresses preference between defeasible rules*

The three special predicates $\mathsf{Assumable}_\neg, \mathsf{Assumable}_\vee, \mathsf{Assumable}_\supset$ provide ways to enable (or attack) arguments that rely on certain proof techniques: proof by contradiction, proof by case distinction and proof of implication introduction.
Furthermore, to facilitate attacks on the validity of rules, we use the naming function n to represent rules in the arguments. Of note is that the rules in $R_\bot$ define the meaning of triviality and should not be attackable. We achieve this by not defining a name for these rules.
An argument can then be constructed inductively by applying one of the rules or one of the special proof techniques on smaller sub-arguments. It should be noted, however, that all rules in $\mathcal{R}$ are treated syntactically and not as logical formulae. For example, it does not automatically hold that $\neg\neg\phi$ is the same as ϕ. Furthermore, we refer to several components of arguments by functions in ASPIC-END: the conclusion of an argument A is denoted as $\mathsf{Conc}(A)$, the assumptions made for the special proofs are denoted as $\mathsf{As}_\neg(A), \mathsf{As}_\vee(A), \mathsf{As}_\supset(A)$, or $\mathsf{As}(A)$ in totality. We call arguments with $\mathsf{As}(A) \neq \emptyset$ as *hypotheticals*. $\mathsf{Sub}(A)$ describes the sub-arguments of A, while $\mathsf{TopRule}(A)$ specifies the last applied rule (if the last step of the argument is indeed a rule application) and $\mathsf{DefRules}(A)$ collects all the defeasible rules in A. With this, we can formally define arguments in a bottom-up manner in ASPIC-END:[7]

Definition 7. *An* argument A *on the basis of an argumentation theory* $\Sigma = (\mathcal{L}, \mathcal{R}, n, <)$ *has one of the following forms:*

1. $A_1, \ldots, A_n \rightsquigarrow \psi$*, where* $A_1, \ldots, A_n$ *are arguments such that there exists an intuitively strict rule* $\mathit{Conc}(A_1), \ldots, \mathit{Conc}(A_n) \rightsquigarrow \psi$ *in* $\mathcal{R}_{is}$*.*
   $\mathit{Conc}(A) := \psi$, $\mathit{As}_\neg(A) := \mathit{As}_\neg(A_1) \cup \cdots \cup \mathit{As}_\neg(A_n)$,
   $\mathit{As}_\vee(A) := \mathit{As}_\vee(A_1) \cup \cdots \cup \mathit{As}_\vee(A_n)$, $\mathit{As}_\supset(A) := \mathit{As}_\supset(A_1) \cup \cdots \cup \mathit{As}_\supset(A_n)$,
   $\mathit{Sub}(A) := \mathit{Sub}(A_1) \cup \cdots \cup \mathit{Sub}(A_n) \cup \{A\}$,
   $\mathit{DefRules}(A) := \mathit{DefRules}(A_1) \cup \cdots \cup \mathit{DefRules}(A_n)$,
   $\mathit{TopRule}(A) := \mathit{Conc}(A_1), \ldots, \mathit{Conc}(A_n) \rightsquigarrow \psi$.

[7] The following definition is verbatim from [CD20] and was provided courtesy of Dr. Marcos Cramer.

2. $A_1, \ldots, A_n \Rightarrow \psi$, *where* $A_1, \ldots, A_n$ *are arguments s.t.* $\mathsf{As}(A_1) \cup \ldots \cup \mathsf{As}(A_n) = \emptyset$ *and there exists a defeasible rule* $\mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \Rightarrow \psi$ *in* $\mathcal{R}_d$.
   $\mathsf{Conc}(A) := \psi$, $\mathsf{As}_{\neg}(A) := \emptyset$,
   $\mathsf{As}_{\vee}(A) := \emptyset$, $\mathsf{As}_{\supset}(A) := \emptyset$,
   $\mathsf{Sub}(A) := \mathsf{Sub}(A_1) \cup \cdots \cup \mathsf{Sub}(A_n) \cup \{A\}$,
   $\mathsf{DefRules}(A) := \mathsf{DefRules}(A_1) \cup \cdots \cup \mathsf{DefRules}(A_n) \cup$
   $\{\mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \Rightarrow \psi\}$,
   $\mathsf{TopRule}(A) := \mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \Rightarrow \psi$.

3. $\mathsf{Assume}_{\neg}(\varphi)$, *where* $\varphi \in \mathcal{L}$.
   $\mathsf{Conc}(A) := \varphi$, $\mathsf{As}_{\neg}(A) := \{\varphi\}$,
   $\mathsf{As}_{\vee}(A) := \emptyset$, $\mathsf{As}_{\supset}(A) := \emptyset$,
   $\mathsf{Sub}(A) := \{\mathsf{Assume}_{\neg}(\varphi)\}$,
   $\mathsf{DefRules}(A) := \emptyset$, $\mathsf{TopRule}(A)$ *is undefined.*

4. $\mathsf{Assume}_{\vee}(\varphi)$, *where* $\varphi \in \mathcal{L}$.
   $\mathsf{Conc}(A) := \varphi$, $\mathsf{As}_{\neg}(A) := \emptyset$,
   $\mathsf{As}_{\vee}(A) := \{\varphi\}$, $\mathsf{As}_{\supset}(A) := \emptyset$,
   $\mathsf{Sub}(A) := \{\mathsf{Assume}_{\vee}(\varphi)\}$,
   $\mathsf{DefRules}(A) := \emptyset$, $\mathsf{TopRule}(A)$ *is undefined.*

5. $\mathsf{Assume}_{\supset}(\varphi)$, *where* $\varphi \in \mathcal{L}$.
   $\mathsf{Conc}(A) := \varphi$, $\mathsf{As}_{\neg}(A) := \emptyset$,
   $\mathsf{As}_{\vee}(A) := \emptyset$, $\mathsf{As}_{\supset}(A) := \{\varphi\}$,
   $\mathsf{Sub}(A) := \{\mathsf{Assume}_{\supset}(\varphi)\}$,
   $\mathsf{DefRules}(A) := \emptyset$, $\mathsf{TopRule}(A)$ *is undefined.*

6. $\mathsf{ProofByContrad}(\neg\varphi, A')$, *where* A' *is an argument such that* $\varphi \in \mathsf{As}_{\neg}(A')$ *and* $\mathsf{Conc}(A') = \bot$.
   $\mathsf{Conc}(A) := \neg\varphi$, $\mathsf{As}_{\neg}(A) := \mathsf{As}_{\neg}(A') \setminus \{\varphi\}$,
   $\mathsf{As}_{\vee}(A) := \mathsf{As}_{\vee}(A')$, $\mathsf{As}_{\supset}(A) := \mathsf{As}_{\supset}(A')$,
   $\mathsf{Sub}(A) := \mathsf{Sub}(A') \cup \{\mathsf{ProofByContrad}(\neg\varphi, A')\}$,
   $\mathsf{DefRules}(A) := \mathsf{DefRules}(A')$, $\mathsf{TopRule}(A)$ *is undefined.*

7. $\mathsf{ReasonByCases}(\psi, A_1, A_2, A_3)$, *where:*
   A_1 *is an argument such that* $\varphi \in \mathsf{As}_{\supset}(A_1)$ *and* $\mathsf{Conc}(A_1) = \psi$,
   A_2 *is an argument such that* $\varphi' \in \mathsf{As}_{\supset}(A_2)$ *and* $\mathsf{Conc}(A_2) = \psi$,
   A_3 *is an argument such that* $\mathsf{Conc}(A_3) = \varphi \vee \varphi'$.
   $\mathsf{Conc}(A) := \psi$,
   $\mathsf{As}_{\neg}(A) := \mathsf{As}_{\neg}(A_1) \cup \mathsf{As}_{\neg}(A_2) \cup \mathsf{As}_{\neg}(A_3)$,
   $\mathsf{As}_{\vee}(A) := (\mathsf{As}_{\vee}(A_1) \setminus \{\varphi\}) \cup (\mathsf{As}_{\vee}(A_2) \setminus \{\varphi'\}) \cup \mathsf{As}_{\vee}(A_3)$,
   $\mathsf{As}_{\supset}(A) := \mathsf{As}_{\supset}(A_1) \cup \mathsf{As}_{\supset}(A_2) \cup \mathsf{As}_{\supset}(A_3)$,
   $\mathsf{Sub}(A) := \mathsf{Sub}(A_1) \cup \mathsf{Sub}(A_2) \cup \mathsf{Sub}(A_3) \cup \{\mathsf{ReasonByCases}(\psi, A_1, A_2, A_3)\}$,
   $\mathsf{DefRules}(A) := \mathsf{DefRules}(A_1) \cup \mathsf{DefRules}(A_2) \cup \mathsf{DefRules}(A_3)$,
   $\mathsf{TopRule}(A)$ *is undefined.*

8. $\supset$-*intro*$(\varphi \supset \psi, A')$, *where* A' *is an argument such that* $\varphi \in$ *As*(A') *and* *Conc*$(A') = \psi$.
Conc$(A) := \varphi \supset \psi$, *As*$_{\neg}(A) :=$ *As*$_{\neg}(A')$,
As$_{\vee}(A) :=$ *As*$_{\vee}(A')$, *As*$_{\supset}(A) :=$ *As*$_{\supset}(A') \setminus \{\varphi\}$,
Sub$(A) :=$ *Sub*$(A') \cup \{\supset$-*intro*$(\varphi \supset \psi, A')\}$,
DefRules$(A) :=$ *DefRules*(A'), *TopRule*(A) *is undefined.*

9. $\forall$-*intro*$(\forall x.\varphi(x), A')$, *where* A' *is an argument such that for some* $x \in \mathcal{L}_v$, *there is no* $\psi \in$ *As*(A') *such that* x *is free in* ψ, *and* *Conc*$(A') = \varphi(x)$.
Conc$(A) := \forall x.\varphi(x)$, *As*$_{\neg}(A) :=$ *As*$_{\neg}(A')$,
As$_{\vee}(A) :=$ *As*$_{\vee}(A')$, *As*$_{\supset}(A) :=$ *As*$_{\supset}(A')$,
Sub$(A) :=$ *Sub*$(A') \cup \{\forall$-*intro*$(\forall x.\varphi(x), A')\}$,
DefRules$(A) :=$ *DefRules*(A'), *TopRule*(A) *is undefined.*

Of note is the construction of hypotheticals, an improvement over the ASPIC^{+} formalism. All three of the proof techniques rely on and work with assumptions: a proof by contradiction for example assumes a fact to then show that this fact leads to $\perp$ and thus (more than) inconsistency. From this we can deduce that the opposite should hold. We can however raise doubt on the applicability of a proof technique (this is one of the ways in which one argument might attack another). Furthermore, we restrict hypotheticals to intuitively strict rules to avoid issues such as proving transposition for defeasible rules ([CD20]).
In ASPIC-END, we distinguish between different types of attacks between arguments (that make up the general $\longrightarrow$ relation). These attacks take the internal structure of the arguments into account.

Definition 8. *For an argumentation theory* $\Sigma = (\mathcal{L}, \mathcal{R}, n, <)$ *we say that an argument* A *attacks an argument* B *(on some* $B' \in$ *Sub*(B)*) if:*

- *Conc*$(A) = \neg\phi$ *(or* $\neg$*Conc*$(A) = \phi$*),* B' *is of the form* $B'_1, ..., B'_n \Rightarrow \phi$ *and* $As(A) = \emptyset$*. This attack is called a* rebuttal*.*

- *Conc*$(A) = \neg n(r)$ *(or* $\neg$*Conc*$(A) = n(r)$*),* *TopRule*$(B) = r$*, there is no* $\phi \in$ *As*(B') *such that* $\neg\phi =$ *Conc*(A') *for* $A' \in$ *Sub*(A) *and there are arguments* $B_1, ..., B_n$ *such that* $B_1 = B'$*,* $B_n = B$*,* $B_i \in$ *Sub*(B_{i+1}) *and* *As*$(A) \subseteq$ *As*$(B_1) \cup ... \cup$ *As*(B_n)*. This attack is called an* undercut*.*

- $As(A) = \emptyset$ *and for a proof method* $m \in \{\neg, \vee, \supset\}$*:* $B' =$ *Assume*$_m(\phi)$ *and* *Conc*$(A) = \neg$*Assumable*$_m(\phi)$*. This is called an* assumption-attack*.*

A rebuttal is an attack on some (possibly intermediary) conclusion of the other argument. Here is one way in which we distinguish between intuitively strict and defeasible rules: if we accept an intuitively strict rule, then we cannot refute its conclusions. The results of a defeasible rule can be rejected more readily. An undercut on the other hand disputes the validity of a rule itself (even an intuitively strict rule) and an assumption attack rejects some assumption that was used for one of the special proof methods.

We have to note, however, that a rebuttal would always cause a symmetric and opposed rebuttal from B' to A. To allow for a "winner" between contradictory arguments, we require that B has the least preferred rule of the arguments.[8]

Definition 9. *Let $\Sigma = (\mathcal{L}, \mathcal{R}, n, <)$ be an argumentation theory and A, B be two arguments from Σ. Then we define the lifting $\prec$ of $<$ to arguments to be the ordering where for any two arguments A, B from Σ, we have $A \prec B$ iff there is $r_b \in \mathsf{DefRules}(B)$ such that $r_b < r_a$ for all rules $r_a \in \mathsf{DefRules}(A)$. Then, we say that A* successfully rebuts *B if A rebuts B on some sub-argument B' and $A \not\prec B$.*

This condition for successful rebuttal is called the "weakest link" of B. with this condition, we can define the induced *attack* relation between arguments.

Definition 10. *Let $\Sigma = (\mathcal{L}, \mathcal{R}, n, <)$ be an argumentation theory and A, B be two arguments from Σ. Then, A successfully attacks B ($A \longrightarrow B$) iff A successfully rebuts, undercuts or assumption-attacks B.*

Given an argumentation theory, we can construct an induced explanatory argumentation framework. For this, we first have to define how one argument might explain another:

Definition 11. *Let A, B be arguments. We say that B* explains *A (on A') iff $A' \in \mathsf{Sub}(A)$, $\mathsf{As}(B) \subseteq \mathsf{As}(A')$ and at least one of the following two cases holds:*

- *$A' \notin \mathsf{Sub}(B)$ and either $A' = (\rightsquigarrow \mathsf{Conc}(B))$ or $A' = (\Rightarrow \mathsf{Conc}(B))$.*
- *$\mathsf{Conc}(B) = n(\mathsf{TopRule}(A'))$ and $\nexists B' \in \mathsf{Sub}(B)$ such that $\mathsf{TopRule}(B') = \mathsf{TopRule}(A')$.*

Thus, we can either explain an argument on one of its conclusions or on why one of its rule is allowed. For general ASPIC-END EAFs, whether an argument sufficiently explains an explanandum is checked with a separate criterion C (a relation on $\mathcal{A} \times \mathcal{X}$ that is provided as part of the definition of the induced EAF). For paradoxical problems (such as the Curry paradox), [CD20] provides a way to turn all paradoxical arguments into explananda that have to be "explained away." We will incorporate this approach in the definition below.
With all of these definitions, we can now present the EAF induced by an ASPIC-END argumentation theory:

Definition 12. *Let $\Sigma = (\mathcal{L}, \mathcal{R}, n, <)$ be an argumentation theory for a paradox. The* explanatory argumentation framework *(EAF)* induced by *Σ is a tuple $(\mathcal{A}, \mathcal{X}, \longrightarrow, \dashrightarrow)$, where:*

- *$\mathcal{A}$ is the set of all arguments that can be constructed from Σ satisfying Definition 7;*
- *$\mathcal{X} = \{E_A \mid A$ is a strict argument (i.e. $\mathsf{DefRules}(A) = \emptyset$) from Σ with $\mathsf{Conc}(A) = \bot$ and $\mathsf{Sub}(A) = \emptyset\}$ is the set of explananda;*

[8] This rule is also called the *weakest link* of B

- $(A, B) \in \longrightarrow$ *iff* A *successfully attacks* B*, where* $A, B \in \mathcal{A}$*;*
- $(A, B) \in \dashrightarrow$ *iff* A *explains* B *according to Definition 11, where* $A, B \in \mathcal{A}$*;*
- $(A, E_B) \in \dashrightarrow$ *iff* A *successfully attacks* B*, where* $A \in \mathcal{A}$ *and* $E_B \in \mathcal{X}$*.*

2.2.3 Introducing Meta-Language to ASPIC-END

As mentioned in section 2.1, the Curry paradox is in part rooted in the nature of natural language logic: while we define the Curry sentence in the object language, we include notions such as references to statements and their truth values, which are more appropriate in the meta-language. While the natural language is powerful enough to pose as its own meta-language, it is not obvious whether the same holds for the arguments that can be defined in ASPIC-END. To model the Curry paradox and arguments about it, we propose to enhance ASPIC-END with meta-language reference.

Definition 13. *Let* $\mathcal{L}$ *be the logical language of an ASPIC-END argumentation theory and* Meta *be a set of predicate symbols not in* $\mathcal{L}$*. Then, the* logical meta-language $\mathcal{L}^{\mathsf{Meta}}$ *is the smallest language such that* $\mathcal{L} \subset \mathcal{L}^{\mathsf{Meta}}$ *and for all* n*-ary predicates* $p \in \mathsf{Meta}$ *and* $l_1, ..., l_n \in \mathcal{L}^{\mathsf{Meta}}$ *we also have* $p(\langle l_1 \rangle, ..., \langle l_n \rangle) \in \mathcal{L}^{\mathsf{Meta}}$*.*

In this construction, we can extend the underlying logical language of an argumentation theory by meta-statements. These consist of new meta-predicates that take other elements of the logical language as arguments. We mark elements of the language with the $\langle\rangle$ brackets. This construction was presented as the “objectification operator” in [CA07][9] and can in fact be likened to *Gödel codes*, which allow representation of logical statements in the logic of arithmetics and are used in the proofs of Gödel's incompleteness theorems.

Since ASPIC-END constructs arguments syntactically, we can extend the logical language for our argumentation theory to include meta-statements without impeding the definitions of ASPIC-END. We will mostly use meta-predicates for the truth or falsity of a statement (represented by the unary $True$ and $False$ respectively).

Another, less well-behaved approach we considered is reasoning about the *arguments* of the argumentation theory itself. This would allow attacks to refer to another argument's structure to justify rejecting it. Furthermore, we might be able to represent rules and argue about their structure in this approach as well. However, such a drastic change might produce unforeseen consequences and we have not seen another work that allows this degree of meta-statement as a reference. In 3.2.2, we have modeled a few chosen counter-arguments to the Curry Paradox that were prevalent in the survey. Some of these arguments could have been presented more concisely with knowledge about the line of reasoning for the Curry paradox or the rules that have been integral to its validity. We will outline on these examples how this approach might benefit reasoning in argumentation theories. We have, however, decided to leave a specific implementation to future work.

[9] Though they refer to an older book by Pollock ([Pol95]) for this operator, where it was used but not explicitly introduced.

3 Practical Results

As stated above, the ASPIC-END framework presents a way to represent formal debate and how contrary positions might counter or justify each other. It is especially useful to represent natural language disputes in a more formalized shape.
In 2018, Cramer and Guillaume performed an (as-of-yet unpublished) survey on the Curry paradox, asking lay people to find counter-arguments to it. More details on the survey will be presented in section 3.1. In this paper, we would like to begin evaluating the responses some participants gave to the survey by identifying some of the more prominent arguments and how they might be modeled in ASPIC-END. To do this, we will first model the standard argument for the Curry paradox (as presented in section 2.1 and in the study) in ASPIC-END, and present the other arguments in reference to this model. Where reasonable, we will add likely counter-arguments to the questionees' arguments to show how these arguments themselves might be countered.
This paper is intended to be as much a proof-of-concept for the ASPIC-END formalism in philosophical debate as it is an evaluation of the survey, giving insight into a lay person's perspective on the Curry paradox. As such, we will have to work with the answers given by lay people to the rather formal Curry paradox. This required us to apply a judicious amount of our own knowledge about the Curry paradox and some interpretation of how some arguments were intended to formulate as ASPIC-END arguments. While we are reasonably certain that we did not miscontrue any argument, we have nonetheless found during the course of this evaluation that with the required amount of interpreting the formalized counter-arguments are as much our own as they are originating from the questionees.

3.1 Survey Methodology

Cramer and Guillaume's survey was conducted from May to June 2018 on 56 university students from both Luxemburg and Germany. They decided to favor students with a background in the MINT field, especially those with prior study on logics and the evaluation of logic formulae. This selection was done so that the questionee were more likely to correctly (and formally) understand the Curry paradox and formulate arguments against it. While this does introduce a selection bias, this survey is not intended to produce generalizable results which the bias would harm the most.
The students were presented with one of two (structurally very similar) formulations of the Curry paradox, one presenting implications as used in the paradox by "if ... then ..." statements and the other by statements of the form "from ... it follows that ...", though we have not focussed on the impacts this might have had on the answers, instead focussing on answers to the first formulation. The survey was conducted in groups of 3 to 6 students, with more than one group at the same time. The survey was divided into 4 steps: students were first provided with a introductory text,outlining the theory behind self-referential natural language logic and the Curry paradox. This included the paradoxes' standard argumentation. They were then asked to individually formulate responses, finding the first step of the standard argumentation that they found

faulty. Afterwards, the students were asked to share and discuss their responses to find a counter-argument that they could agree on as a group. The discussions were recorded in audio and as a transcript. Lastly, the students were asked to provide another individual answer in light of the discussion.
In an earlier study with the same structure, Cramer and Guillaume motivated the group discussion methodology of the survey as follows:[10] (see [CG19])

> Previous results showed that individual performance, which has generally been reported to be quite poor in pure logic and reasoning tasks, could actually be enhanced by cooperative discussion with peers. For instance, faced with the Wason selection task [Was66], humans solving the task in groups achieved a level of insight that was qualitatively superior to the one achieved by single individuals [Gei98, Aug08]. Additionally, and more generally, discussion with peers was shown to substantially improve motivation to solve a given task [PSB$^+$95]. For these reasons, we decided to incorporate in our methodology a cooperative discussion to help participants to elaborate and enrich their thinking. This collective step with peers was designed to obtain an evaluation of the justification status more reliable than a single individual judgment. Such reliability is crucial to test the cognitive plausibility of our predictions.

3.2 Modeling the Survey

In the following sections, we will present an ASPIC-END argumentation theory modeling some of the survey results. Our main focus for this will be the relevant rules needed to construct arguments and counter-arguments. When appropriate, we will present abstract rule schemes, but we will for the most part only present the necessary and instantiated rules. There is an overview of the constructed argumentation theory in section 3.3.

3.2.1 The Argument for the Curry paradox

In the survey, the questionees were given a short introduction into natural language logic sentences and the Curry sentence[11] in particular. Furthermore, they were given a few logical rules for use in deriving the Curry sentence: *True-Introduction*, *True-Elimination* and *If-Elimination*, also called *modus ponens*.
All logical rules were given as logical sentences in natural language and as a derivation figure for *natural deduction*. This requires us to distinguish between a given sentence holding, and the same sentence being assigned the value true as a meta-statement in the natural language.

- *True-Introduction*: "If S, then 'the sentence S is true' "

[10] The citation and references are provided courtesy of Dr. Marcos Cramer.

[11] We are referring to the standard Curry sentence in this survey, i.e. "If this sentence is true, then the moon is made of cheese."

- *True-Elimination*: "If 'the sentence S is true', then S"
- *If-Elimination*: "If S and 'If S then T', then T"

Furthermore, the questionees were introduced to the proof scheme for natural implication: an implication "If S then T" holds if we can show T when we assume that S holds. This rule is one of the special proof schemes in ASPIC-END and we do not need to produce a separate rule for it.
For the other rules as well as some miscellaneous logical properties, we propose the following rules in ASPIC-END (see table 2)

$r_{T\_El}(\varphi)$	$True(\langle\varphi\rangle)$	$\rightsquigarrow \varphi$	True-Elimination
$r_{T\_In}(\varphi)$	φ	$\rightsquigarrow True(\langle\varphi\rangle)$	True-Introduction
$r_{if\_El}(\varphi_1, \varphi_2)$	$\varphi_1 \supset \varphi_2, \varphi_1 \rightsquigarrow$	φ_2	If-Elimination (i.e. modus ponens)
$r_{F\_T}(\varphi)$	$False(\langle\varphi\rangle)$	$\rightsquigarrow \neg True(\langle\varphi\rangle)$	opposite truth values
$r_{T\_F}(\varphi)$	$True(\langle\varphi\rangle)$	$\rightsquigarrow \neg False(\langle\varphi\rangle)$	opposite truth values
$r_{contra}(\varphi)$	$True(\langle\varphi\rangle), False(\langle\varphi\rangle)$	$\rightsquigarrow \bot$	opposite truth values
$r_{TF\_cases}(\varphi)$		$\rightsquigarrow True(\langle\varphi\rangle) \vee False(\langle\varphi\rangle)$	opposite truth values

Table 2: Rules of Logic for the Curry Paradox

Additionally, we can represent the Curry sentence in ASPIC-END in a very simple form. For this, we will use a nullary function symbol c for the sentence itself,[12] and a nullary relation symbol M for the statement "the moon is made of cheese." Furthermore, to have the standard argument be an explananda for the induced paradox EAF, we require that any Curry sentence leads to triviality.

Rule name	Rule		Justification
r_{Def1}	$True(c) \rightsquigarrow$	$True(\langle True(c) \supset M\rangle)$	Definition of c (the Curry sentence)
r_{Def2}	$True(\langle True(c) \supset M\rangle) \rightsquigarrow$	$True(c)$	Definition of c (the Curry sentence)
r_{triv}	$True(c), M \Rightarrow$	$\bot$	Existence of a Curry sentence leads to triviality

Table 3: The Curry Sentence

With these constructions, we can formulate an argument for the Curry sentence in table 4 that is structurally similar to the argument presented in the survey (in fact, it is structurally identical except for showing natural implication via a proof scheme instead of a rule):

[12] This c represents a logical formula and is thus only used with $True$.

Argument			Top Rule
$A_1 =$		$Assume_{\supset}(True(c))$	
$A_2 =$	$A_1 \rightsquigarrow$	$True(\langle True(c) \supset M\rangle)$	r_{Def1}
$A_3 =$	$A_2 \rightsquigarrow$	$True(c) \supset M$	$r_{T\_El}(True(c) \supset M)$
$A_4 =$	$A_3, A_1 \rightsquigarrow$	M	$r_{if\_El}(True(c), M)$
$B_1 =$		$\supset -intro(True(c) \supset M, A_4)$	
$B_2 =$	$B_1 \rightsquigarrow$	$True(\langle True(c) \supset M\rangle)$	$r_{T\_In}(True(c) \supset M)$
$B_3 =$	$B_2 \rightsquigarrow$	$\langle True(c)\rangle$	r_{Def2}
$B_4 =$	$B_1, B_3 \rightsquigarrow$	M	$r_{if\_El}(True(c), M)$
$B_5 =$	$B_1, B_4 \Rightarrow$	$\perp$	r_{triv}

Table 4: The Standard Curry argument

3.2.2 The Responses

From the responses given to the survey, we have picked four that were convincing to the group that they were developed in (this includes some group answers) or that presented an interesting train of thought to be modeled in ASPIC-END. In the following, we will present these four counter-arguments to the Curry argument, and discuss flaws / possible angles of attack these counter-arguments might have. More so than the argument for the Curry sentence, representing these counter-arguments in ASPIC-END required some degree of interpretation of the questionee's answers. While the Curry sentence was presented in the survey in a written, but highly formalized manner, from which the ASPIC-END representation followed almost directly, the answers were more in need of an interpretation: some answers left out or only hinted at minor or obvious logical steps. We have tried to preserve the central ideas for every argument here.

Approach 1: Assigning Truth Values "Makes No Sense"

This approach states that assigning binary truth values to the Curry sentence leads to quite problematic consequences. Assuming that C is true leads to an arbitrary result. For a different Curry sentence C', this result may even be a contradiction, leading us to concluding that the logic is inconsistent. On the other hand, assuming that it is false is contradictory, as it is the antecedent of itself, and thus would make the implication true by definition of an implication. As neither of the truth values is adequate for C, the questionees argued that one should not assign a truth value to the sentence (especially since it is self-referencing). This is a *truth-value gap* explanation (this type of explanation is used e.g. for the Liar's paradox).

Name	Rule	Justification
$r_{Def1'}$	$True(c') \rightsquigarrow True(\langle True(c') \supset False(c')\rangle)$	Definition of a contradictory Curry variant
$r_{Def2'}$	$True(\langle True(c') \supset False(c')\rangle) \rightsquigarrow True(c')$	Definition of a contradictory Curry variant
$r_{if\_El'}$	$True(c') \supset False(c'), True(c') \rightsquigarrow False(c')$	If-Elimination (i.e. modus ponens)
r_{false_ante}	$\neg True(c') \rightsquigarrow True(c') \supset False(c)$	False Antecedent allows any conclusion
	$\rightsquigarrow sRef(c')$	c' is a self-referencing sentence
	$\Rightarrow similar(c', c)$	Both Curry sentences have a similar structure
	$similar(c', c), tGap(c') \Rightarrow tGap(c)$	If c' belongs to the truth gap, then so does c
	$\neg(True(c') \lor False(c')), sRef(c') \Rightarrow tGap(c')$	Contradictory self-referencing sentences belong in the truth value gap
	$tGap(c) \Rightarrow \neg\mathsf{Assumable}_{\supset}(c)$	A sentence assigned to the truth value gap should not have a truth value

Table 5: Rules for the Truth-Value Gap Approach

They effectively argue that for a contradictory variant C' of the Curry sentence, deriving a contradiction shows that it should be consigned to the truth-value gap ("$tGap(\cdot)$"), where we cannot assign any traditional truth value to the sentence. This includes assuming a truth value to show the natural implication in the $\supset$-introduction. Additionally, they argue that since both Curry sentences are structurally very similar ("$similar(\cdot, \cdot)$"), the same can be said for our original Curry sentence.

Argument			Top Rule
$C_1 =$		$Assume_{\neg}(True(c') \vee False(c'))$	
$D_1 =$		$Assume_{\vee}(True(c'))$	
$D_2 =$	$D_1 \rightsquigarrow$	$True(\langle True(c') \supset False(c')\rangle)$	$r_{Def1'}$
$D_3 =$	$D_2 \rightsquigarrow$	$True(c') \supset False(c')$	$r_{T\_El}(True(c') \supset False(c'))$
$D_4 =$	$D_1, D_3 \rightsquigarrow$	$False(c')$	$r_{if\_El'}$
$D_5 =$	$D_1, D_4 \rightsquigarrow$	$\perp$	$r_{contra}(c')$
$E_1 =$		$Assume_{\vee}(False(c'))$	
$E_2 =$	$E_1 \rightsquigarrow$	$\neg True(c')$	
$E_3 =$	$E_2 \rightsquigarrow$	$True(c') \supset False(c')$	r_{false_ante}
$E_4 =$	$E_3 \rightsquigarrow$	$True(\langle True(c') \supset False(c')\rangle)$	$r_{T\_In}(True(c') \supset False(c'))$
$E_5 =$	$E_4 \rightsquigarrow$	$True(c')$	$r_{Def2'}$
$E_6 =$	$E_1, E_5 \rightsquigarrow$	$\perp$	$r_{contra}(c')$
$C_2 =$		$ReasonByCases(\perp, D_5, E_6, C_1)$	
$C_3 =$		$ProofByContrad(\neg(True(c') \vee False(c')), C_2)$	
$C_4 =$	$\Rightarrow$	$sRef(c')$	
$C_5 =$	$C_3, C_4 \Rightarrow$	$tGap(c')$	
$C_6 =$	$\Rightarrow$	$similar(c', c)$	
$C_7 =$	$C_6, C_5 \Rightarrow$	$tGap(c)$	
$C_8 =$	$C_7 \Rightarrow$	$\neg \mathsf{Assumable}_{\supset}(c)$	

Table 6: Truth-Value Gap Argument

Approach 2: Two Kinds of Truth Assertions

The idea stated in these answers is that the argumentation should not take truth assertions on the meta-level[13] and actual truth in a statement to be the same thing. They further argue that step 4 (the application of modus ponens) is the first to mix these two types of truth values by concluding from a material implication and the meta-truth assertion of its antecedent that its actual consequence should hold.

To model this argument in ASPIC, we will require an alternative way of modelling the Curry paradox. Instead of using one $True$ predicate, we will distinguish between meta-level truth assertions $True_m$ and internal truth values $True_i$. In the new modeling, we will use the internal truth value whenever it is first introduced inside a different truth statement. Outside of these sentences, statements about truth values are assumed to be meta-level assertions.

This alternative model of the Curry sentence will use its own rules, which are derived from the way the original rules are used (and consequently, the truth types they use) in the standard argument. These rules are "disambiguated" versions of the original rules (and we use a $disamb(\cdot)$ function to represent this).[14] Ultimately, we will show that one of these disambiguated rules (the modus ponens) is not actually an acceptable version of the modus ponens and must therefore be rejected.

[13] Statements talking about the truth of other statements

[14] The rules provided below are not the complete disambiguated rule corpus, but just an illustrative smaller portion.

Name	Rule	Justification
$r_{Def1''}$	$True_m(c) \rightsquigarrow True_m(\langle True_i(c) \supset M \rangle)$	Alternative definition of c
$r_{T\_El''}$	$True_m(\langle True_i(c) \supset M \rangle) \rightsquigarrow True_i(c) \supset M$	True-Elimination
$r_{dis\_T\_El}(\varphi)$	$\Rightarrow n(disambig(r_{T\_El}(True(c) \supset M))) = n(r_{T\_El''})$	Disambiguated True-Elimination
$r_{MP''}$	$True_i(c), True_i(c) \supset M \rightsquigarrow M$	Real Modus Ponens
r_{dis_MP}	$\Rightarrow n(disambig(r_{if\_El})) = n(True_m(c), True_i(c) \supset M \rightsquigarrow M)$	Disambig Modus Ponens

Table 7: Disambiguated Curry Rules

For this construction we require two additional features beyond ASPIC-END's normal capabilities. One is an equality symbol (this is already used above). We will require the equality symbol to adhere to the defining properties of equality: it is an equivalence relation[15] where equal object can additionally be replaced by one another in all functions and relations. This construction is easily done for cases with a fixed set of relations and functions. While it may seem as if we did not have the same fixed set of relations and functions for all arguments in ASPIC-END (as they all utilise different relation symbols), in theory this is not the case. Theoretically, all arguments provided in this draft are not so much extensions of an argumentation theory as they describe parts of one joint argumentation theory (that has all the relations and functions in its $\mathcal{L}$).
The other feature is a relaxation of the function naming n. To argue about non-rules, we need to extend n to rule-like constructs that are not in the rule sets of our argumentation theory. We would like to additionally provide a predicate $rule(\cdot)$ which contains all actual rules of the theory. Then, we can construct the following rule (let $\tau = True_m(c), True_i(c) \supset M \rightsquigarrow M$ in the following):

Name	Rule	Justification
r_{rule_MP}	$\Rightarrow \quad \neg rule(n(\tau))$	By Definition of $rule(\cdot)$
r_{nprop_MP}	$\neg rule(n(\tau)), n(disambig(r_{if\_El})) = n(\tau) \Rightarrow \quad \neg n(r_{if\_El})$	The disambiguation is not a proper rule

Table 8: Disambiguation Argument

It should be noted that this argument is one of the cases where reasoning about arguments (as proposed in section 2.2.3) in ASPIC-END could prove useful: The distinction between internal and meta-level truth assertions is not motivated or explained by the argument, instead we just provided a differentiated version of the rules. Thus, no counter-argument can be made against the construction process as such. Furthermore,

[15] I.e. A symmetric, reflexive and transitive relation.

the construction is based on the way the truth assertions occurred in the standard argument. The construction process could be presented in a more integrated manner if we could access the standard argument from within the argumentation theory.

Possible Counter-Argument:
In this argument, most of the actual argumentation is baked into the construction of the specific rules, from which the remaining argument can be constructed straight-forwardly. We have found that in the disambiguation of the used modus ponens both the internal and the meta-truth occur, whereas we would expect an actual variant of the modus ponens to use the same truth type twice.

Argument			Top Rule
$F_1 =$	$\Rightarrow$	$n(disambig(r_{if\_El})) = n(\tau)$	r_{dis_MP}
$F_2 =$	$\Rightarrow$	$\neg rule(n(\tau))$	r_{rule_MP}
$F_3 =$	$F_2, F_1 \Rightarrow$	$\neg n(r_{if\_El})$	r_{nprop_MP}

Table 9: Rules for Truth-Variant Coincidence

While this is a rather interesting type of counter-argument to the Curry sentence, it is nonetheless rather problematic to state that there are two independent types of truth values. The given construction and the subsequent argument hinge on there being a strong enough distinction between internal and meta-truth that we cannot conclude that $True_i(c)$ holds whenever $True_m(c)$ holds (and vice versa). However, while the construction makes these two predicates syntactically different, they are both disambiguations of the same truth predicate and should be expected to hold at the same times. This is especially important when looking at the disambiguated rule for modus ponens: if we assume both truth types to only hold concurrently, then the disambiguated MP is an implied rule of the knowledge base. (For the following, let $a_1 = True_i(c) \wedge True_i(c) \supset M$ and $a_2 = True_m(c) \wedge True_i(c) \supset M$ be the antecedents of the two MP variants).

Name	Rule	Justification
$r_{iM\_Conc1}(\varphi)$	$\Rightarrow \ True_i(\varphi) \supset True_m(\varphi)$	Concurrent truth types
$r_{iM\_Conc2}(\varphi)$	$\Rightarrow \ True_m(\varphi) \supset True_i(\varphi)$	Concurrent truth types
r_{iM_proj}	$a_2 \Rightarrow \ True_m(c)$	Projection of a conjunction
r_{iM_repl}	$a_2, True_i(c) \Rightarrow \ a_1$	Extending the conjunction
r_{quasi_Rule}	$a_2 \supset M \Rightarrow \ \neg\neg rule(n(\tau))$	τ is an implied rule

Table 10: Rules for the Disambiguation Counter-Argument

With this, we can construct an attack on the counter-argument rather straightforwardly by showing that τ is an implied rule of the argumentation theory.

Argument	Top Rule
$F'_1 = \quad Assume_{\supset}(a_2)$	
$F'_2 = \quad F'_1 \Rightarrow \ True_m(c)$	r_{iM_proj}
$F'_3 = \quad \Rightarrow \ True_m(c) \supset True_i(c)$	$r_{iM\_Conc2}(c)$
$F'_4 = \quad F'_1, F'_3 \Rightarrow \ True_i(c)$	$r_{if\_El}(True_m(c), True_i(c))$
$F'_5 = \quad F'_1, F'_4 \Rightarrow \ M$	$r_{MP''}$
$F'_6 = \quad \supset -Intro(a_2 \supset M, F'_5)$	
$F'_7 = \quad F'_6 \Rightarrow \ \neg\neg rule(n(\tau))$	r_{quasi_Rule}

Table 11: The Disambiguation Counter-Argument

Approach 3: The Truth Predicate may not be Definable
This approach is based on the fact that C may not actually be a decidable sentence, since it uses the *True* predicate. A questionee provided the outline for a proof that the *True* predicate is not calculatable in some logic theories (prime among them arithmetics[16]) according to Tarski's Undefinability theorem. This provides an example of a logic theory where truth assertions in the object language (i.e. the language reasoned over) are problematic. Without any concrete evidence to the contrary, this *could* possibly also hold in the object language that C is formulated in.

[16] The theory where addition and multiplication of natural numbers can be constructed

Rule name	Rule	Justification
r_{a_Tarski}	$\rightsquigarrow UndefTheo(arith)$	Tarski's Undefinability Theorem holds for arithmetics
r_{arith_Undef}	$UndefTheo(arith) \rightsquigarrow T\_Undef(arith)$	Truth not definable in arithmetics
r_{maybe_Undef}	$T\_Undef(arith) \Rightarrow maybe(\langle T\_Undef(natural\_L)\rangle)$	If there is a logic where truth is undefinable, then this *might* hold for this object language as well
r_{nat_Undef}	$maybe(\langle T\_Undef(natural\_L)\rangle) \Rightarrow T\_Undef(natural\_L)$	Maybe truth is undefinable
r_{c_Desc}	$\Rightarrow contains\_T(r_{Def1})$	Description of C contains truth predicate
r_{c_contra}	$T\_Undef(natural\_L), contains\_T(r_{Def1}) \Rightarrow \neg n(r_{Def1})$	Description of C may not be well-defined

Table 12: Rules for the Truth-Undefinability Argument

The argument against the Curry sentence can now be constructed very straightforwardly:

Argument			Top Rule
$G_1 =$	$\rightsquigarrow$	$UndefTheo(arith)$	r_{a_Tarski}
$G_2 =$	$G_1 \rightsquigarrow$	$T\_Undef(arith)$	r_{arith_Undef}
$G_3 =$	$G_2 \Rightarrow$	$maybe(\langle T\_Undef(natural\_L)\rangle)$	r_{maybe_Undef}
$G_4 =$	$G_3 \Rightarrow$	$T\_Undef(natural\_L)$	r_{nat_Undef}
$G_5 =$	$\Rightarrow$	$contains\_T(r_{Def1})$	r_{c_Desc}
$G_6 =$	$G_3, G_5 \Rightarrow$	$\neg n(r_{Def1})$	r_{c_contra}

Table 13: The Truth-Undefinability Argument

Possible Counter-Argument:
There is a major flaw in this approach: from showing that $True$ is undefinable in arithmetic logic we can only conclude that the same might hold for natural language. We instead take a leap of logic and state that $True$ cannot be definable in natural language. This leap is not substantiated in any way and should thus not be taken.

Rule name	Rule	Justification
$r_{leap}(r)$	$leap(r) \Rightarrow \neg n(r)$	Leaps of logic should not be taken
r_{leap_natL}	$\Rightarrow leap(r_{nat\_Undef})$	From "maybe" to "definitely" is a leap of logic

Table 14: Rules for the Truth-Undefinability Counter-Argument

The attack against this argument is then very straightforward.

Approach 4: C "is" Infinite

This approach is significant insofar as different groups of questionees have arrived at it as a counter to the Curry sentence. The main thrust of this approach is that any interpretation of the Curry sentence must by its self-referencing nature be "infinite" in the sense that any derivation of it in the proof calculus of natural deduction would repeatedly try to apply the definition of C to its body and only extend the Curry sentence without halting.[17] Thus, the questionees argued that C cannot be reduced to a truth value.

Rule name	Rule	Justification
r_{circ_Def}	$\Rightarrow \; Circ\_Def(c)$	c has a circular definition
$r_{finitary}$	$Circ\_Def(c) \Rightarrow \; \neg Finitary(c)$	The interpretation of c must be infinite
r_{fin_only}	$\Rightarrow \; Fin\_Only$	We allow only finitary sentences
$r_{fintary\_contra}$	$\neg Finitary(c), Fin\_Only \Rightarrow \; \neg n(r_{Def1})$	A non-finitary sentence is rejected

Table 15: Rules for the "Infinite Derivation" Argument

The attack against the standard argumentation is then very straightforward:

Argument	Top Rule
$H_1 = \; \Rightarrow \; Circ\_Def(c)$	r_{circ_Def}
$H_2 = \; H_1 \Rightarrow \; \neg Finitary(c)$	$r_{finitary}$
$H_3 = \; \Rightarrow \; Fin\_Only$	r_{fin_only}
$H_4 = \; H_2, H_3 \Rightarrow \; \neg n(r_{Def1})$	$r_{fintary\_contra}$

Table 16: The "Infinite Derivation" Argument

Possible Counter-Argument:

While this argument asserts that C does not have a finitary derivation (and that is true), it should still be possible to present an interpretation (i.e. a truth-value assignment) for C and check the consistency of this interpretation (though perhaps not in a finitary manner in the calculus of natural deduction). In fact, restricting C to only finitary derivation in this specific calculus (i.e. the Fin_Only predicate) is an unnecessary restriction that is not motivated as such. We are free to reject this undue restriction (and by doing this, rebut the Fin_Only restriction). In that case, which rebuttal is successful is decided by the preference relation $<$ that is part of the argumentation theory.

[17] This type of behaviour where a sentence cannot be derived in a finite number of steps is called *non-finitary*.

3.3 The Constructed Argumentation Theory

In section 3.2, we presented the standard argumentation for the Curry paradox and 4 arguments to counter the Curry paradox. While we do not intend to rehash the rule sets and argumentations, we would like to provide an overview of the developed arguments and counter-arguments. Additionally, we can construct the AC and EC extensions of the given arguments from this.

In section 3.2, the standard argumentation was formalized by the argument B_5 (and its sub-arguments $A_1...A_4$ and $B_1...B_4$). Similarly, the arguments and counter-arguments were assigned their own unique letters, distinguishing self-contained sub-arguments by individual letters where appropriate. In table 1, we have provided a short list of the arguments and their main concepts.

Argument		Main Idea
Truth-Value Gap	C, D, E	The Curry sentence should not be asigned a truth value.
Two Kinds of Truth Values	F	The standard argument improperly mixes meta-level and internal truth assertions
	F'	Both types of truth values should coincide
Truth Predicate may be Undefinable	G	Truth Predicate is undefinable for arithmetic logic (Tarski's Undefinability Theorem)
	G'	Tarski's Theorem may or may not have an influence on natural language logic
The Curry sentence is infinite	H	Resolution of the Curry sentence does not terminate
	H'	Only finitary approaches reject a non-terminating sentence when there are others semantics possible

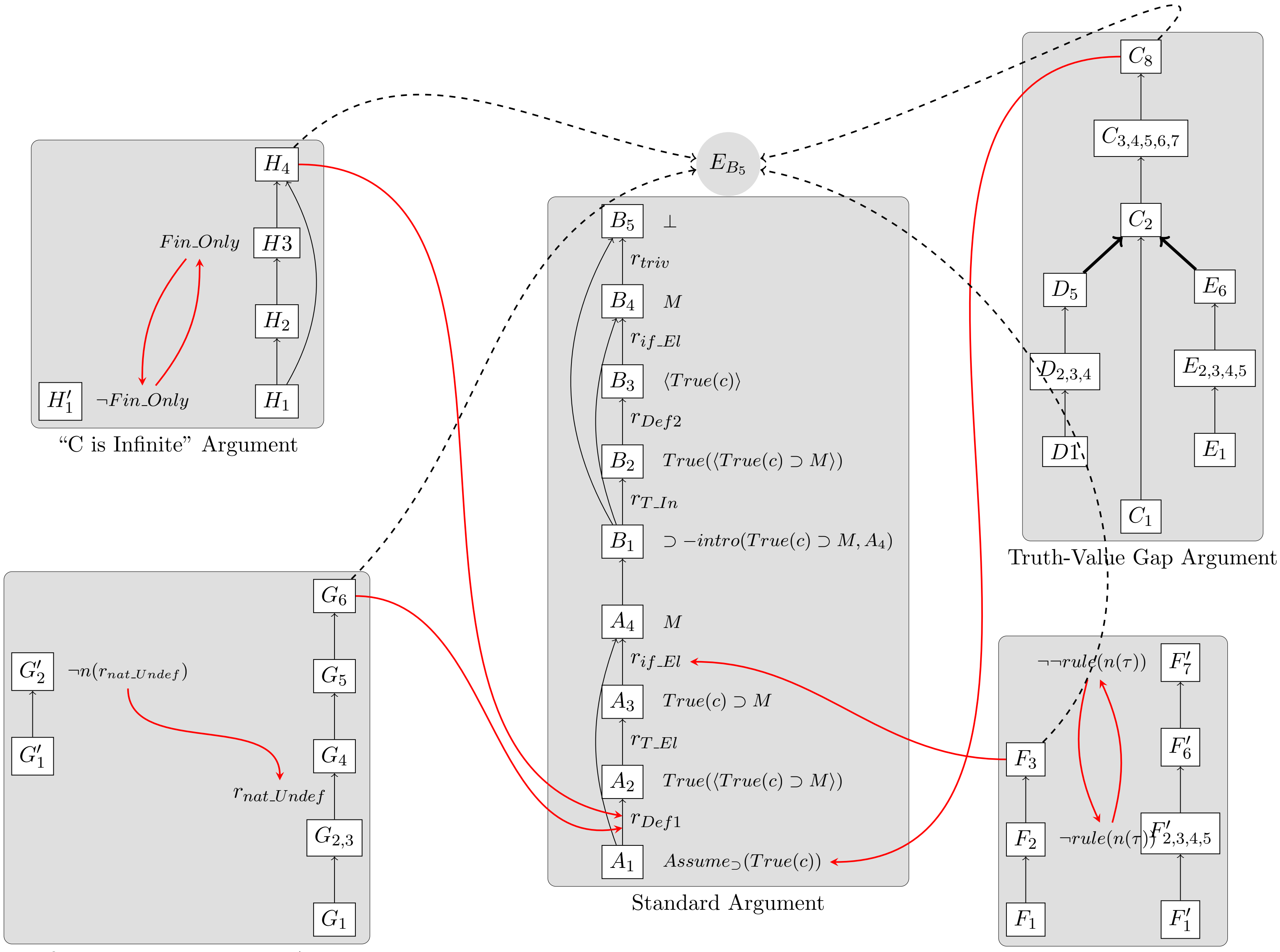
Figure 1: Modeled Arguments against the Curry Paradox

In figure 2, we have presented the most important connections between the five main arguments. Every argument is connected with its direct sub-arguments (and thus, indirectly linked with all of its sub-arguments) and the direct *attack* relation between arguments is denoted with red arrows. It should be noted that all arguments that rely on an attacked rule or subargument are in turn indirectly attacked as well.

As can be seen from the diagram, there are mutual attacks between F_3 and F'_7 as well as H_4 and H'_1. Depending on the choice of $<$, one or the other may be in an admissible extension of the EAF. We will not provide a $<$ here, since this preference relation would represent a philosophical standpoint and this paper is not intended to to provide its own conclusions on the Curry paradox. In this case, $<$ falls back to its default value, the empty ordering.

In the resulting EAF, the only paradox explanandum is E_{B_5}, with C_8, F_3, G_6 and H_4 explaining this explanandum. The *explanation* relation is denoted by the dashed arrows in the diagram. (As a note, with only one explanandum, $<_p$ is a trivial relation.)

Consequently, there are four *argumentative core extensions* of the EAF. Each contain C_8, D_6 and E_5 as well as all of their subarguments. Since the G_6 argument is attacked and cannot be defended, it is not in any of the AC extensions. Additionally, the extensions contain two of the mutually-attacking F_3/F'_7 and H_4/H'_1 to be set-maximal and consistent. The *explanatory core extensions* are $\{C_8\}$, $\{F_3, F_2\}$ and $\{H_4, H_3\}$ (even the set-minimal EC extensions have to contain F_2 as well as H_3 to defend against their respective counter-arguments).



“C is Infinite” Argument

Truth-Value Gap Argument

Standard Argument

4 Conclusion

In this paper, we set out to present and model some lay answers to the problem that the Curry paradox poses. We have managed – with varying levels of elaborateness – to construct formalized ASPIC-END arguments for the responses. In section 3.3, we roughly sketched the resulting argumentation theory and identified arguments that can be considered justified (as elements of the AC and EC extensions). Furthermore, argumentation frameworks have proven to be an useful tool to represent such formalized topics as they are presented in philosophical logic. We have also found ways to represent meta-language references in ASPIC-END, motivated by the blurring between meta-language and object language in natural language logic.

Ultimately we have however found that formalizing the arguments of lay people required (sometimes rather liberal) interpretation of their answers. It is questionable how close the arguments are to what the questionees originally intended. While the collected survey data sometimes helped clarify points of contention, there are still instances that (while internally consistent and following the same line of reasoning) may deviate from the original ideas of the participants.

Two areas that we find to be more promising for future research are to look into (1) formalizing more concise (possibly academic) argumentations and (2) meta-language references for argumentation frameworks. While some of our arguments may unfortunately deviate from the original ideas the participants sought to express, this problem should be alleviated in more precise and academic discussions. There, argumentation frameworks, especially ASPIC-END, may become a tool to formalize debates and maybe even construct new arguments for this debate. It might also be interesting to further study the effect different levels of meta-language statements might have on the expressiveness and feasibility of argumentation theories.

References

[Aug08] Maria Augustinova. Falsification cueing in collective reasoning: example of the Wason selection task. *European Journal of Social Psychology*, 38(5):770–785, 2008.

[BM13] JC Beall and Julien Murzi. Two flavors of curry's paradox. *The Journal of Philosophy*, 110(3):143–165, 2013.

[CA07] Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.

[CD20] Marcos Cramer and Jérémie Dauphin. A structured argumentation framework for modeling debates in the formal sciences. *Journal for General Philosophy of Science*, 51(2):219–241, 2020.

[CG19] Marcos Cramer and Mathieu Guillaume. Empirical study on human evaluation of complex argumentation frameworks. In *European Conference on Logics in Artificial Intelligence*, pages 102–115. Springer, 2019.

[Dun95] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence*, 77(2):321–357, 1995.

[Gei98] David Moshman Molly Geil. Collaborative Reasoning: Evidence for Collective Rationality. *Thinking & Reasoning*, 4(3):231–248, 1998.

[Hod18] Wilfrid Hodges. Tarski's Truth Definitions. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2018 edition, 2018.

[MP14] Sanjay Modgil and Henry Prakken. The aspic+ framework for structured argumentation: a tutorial. *Argument & Computation*, 5(1):31–62, 2014.

[Pol95] John L Pollock. *Cognitive carpentry: A blueprint for how to build a person.* Mit Press, 1995.

[PSB$^+$95] J. Piaget, L. Smith, T. Brown, R. Campbell, N. Emler, and D.N.M. Ferrari. *Sociological Studies*. Routledge, 1995.

[SB18] Lionel Shapiro and Jc Beall. Curry's Paradox. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2018 edition, 2018.

[ŠS13] Dunja Šešelja and Christian Straßer. Abstract argumentation and explanation applied to scientific debates. *Synthese*, 190(12):2195–2217, 2013.

[SW20] Yaroslav Shramko and Heinrich Wansing. Truth Values. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2020 edition, 2020.

[T$^+$56] Alfred Tarski et al. The concept of truth in formalized languages. *Logic, semantics, metamathematics*, 2(152-278):7, 1956.

[Was66] Peter C. Wason. Reasoning. In B. Foss, editor, *New Horizons in Psychology*, pages 135–151. Harmondsworth: Penguin Books, 1966.