Consistent Rational Argumentation (in Politics)

Prof. Dr. Christoph Benzmüller and Dr. Alexander Steen
Freie Universität Berlin and University of Luxembourg

Formal Analysis of Arguments in Politics and Beyond

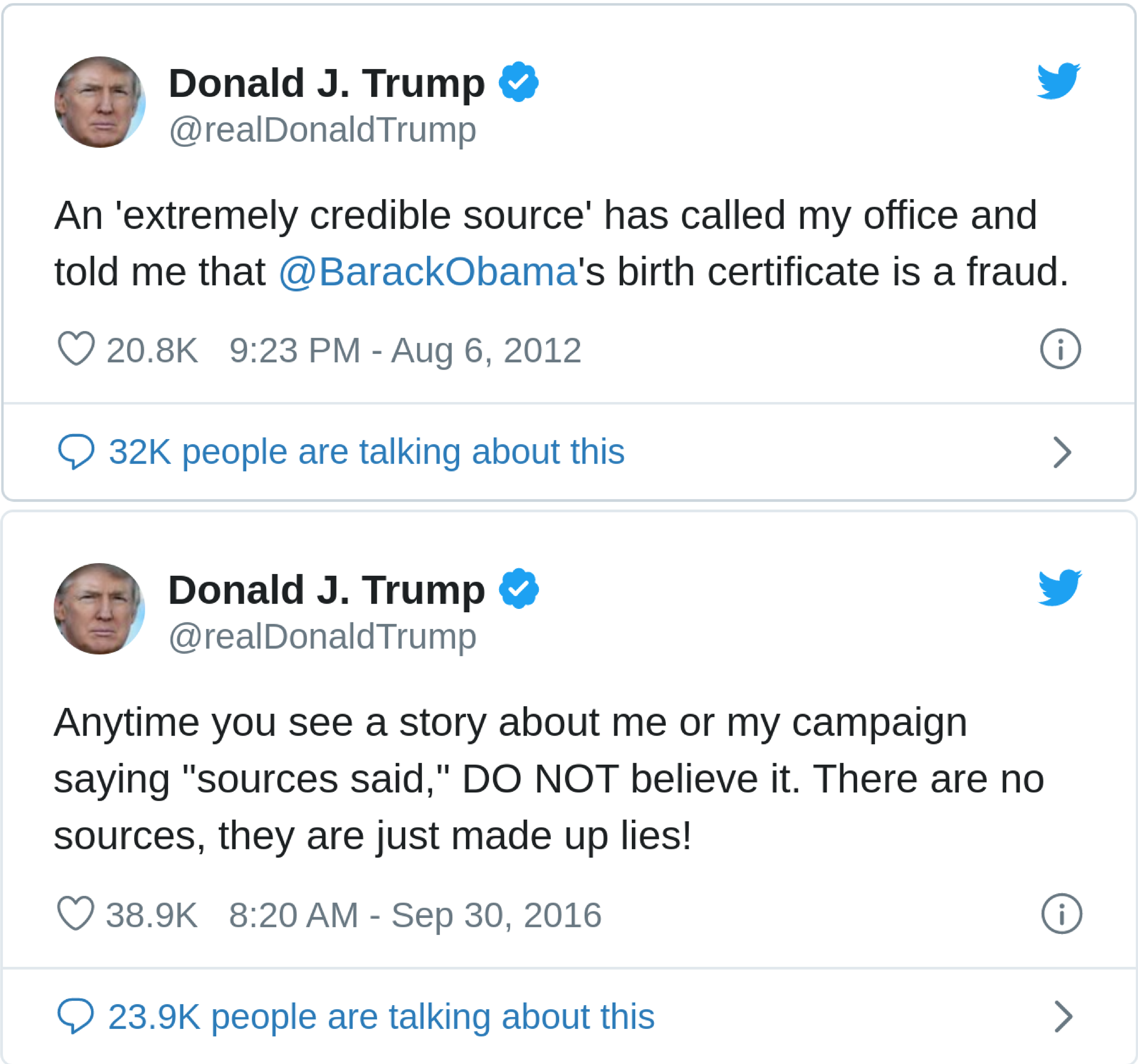
Vision: Automated detection of irrational and flawed arguments.

Hypothesis: The universal meta-logical reasoning approach [1] (utilising higher-order logic) can be adapted to support the task.

Challenges/Questions:

- Different nature of political & legal arguments?
- Which modalities play a role and need to be combined?
- Do reasoning systems (e.g. Leo-III [3]) still perform well?
- Combination with other NLP tools?
- General methodology? — Computational hermeneutics [2]

Example tweets:



Donald J. Trump @realDonaldTrump

An 'extremely credible source' has called my office and told me that @BarackObama's birth certificate is a fraud.

20.8K 9:23 PM - Aug 6, 2012

32K people are talking about this

Donald J. Trump @realDonaldTrump

Anytime you see a story about me or my campaign saying "sources said," DO NOT believe it. There are no sources, they are just made up lies!

38.9K 8:20 AM - Sep 30, 2016

23.9K people are talking about this

Figure 1: Tweets of POTUS Donald Trump.

Question: Can the irrationality be formally detected?

Argument Preprocessing

The NL text is converted into computer-readible data structure that incorporates every word's position and function within the sentence. As an example, the Stanford NLP parser produces the following output for the first tweet from Fig. 1:

```
(ROOT
  (S
    (NP (DT An)
      (NP (`` `)
        (ADJP (RB extremely) (VBN credible))
        (NN source) ('' ')))
    (VP
      (VP (VBZ has)
        (VP (VBN called)
          (NP (PRP my) (NN office))))
      (CC and)
      (VP (VBD told)
        (NP (PRP me))
        (SBAR (IN that)
          (S
            (NP
              (NP (NNP BarackObama) (POS 's))
              (NN birth) (NN certificate))
            (VP (VBZ is)
              (NP (DT a) (NN fraud)))))))
    (. .)))
```

→

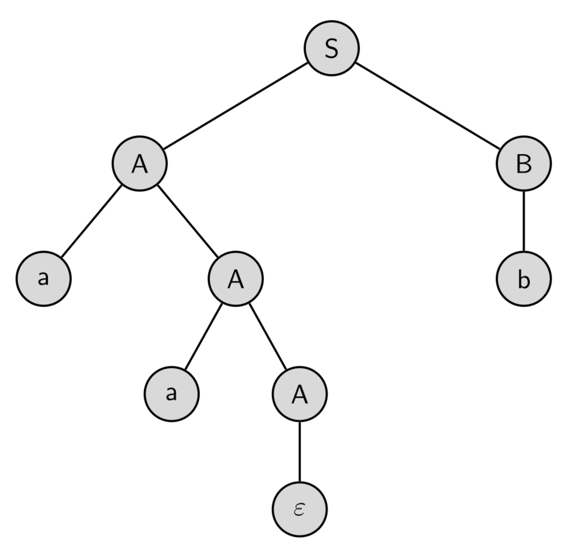


Figure 2: Pre-processing out with semantical NLP parser (left), transformation into computer-readible data structure for further processing (right).

References

[1] C. Benzmüller.
Universal (meta-)logical reasoning: Recent successes.
Science of Computer Programming, 172:48–62, 2019.

[2] D. Fuenmayor and C. Benzmüller.
A computational-hermeneutic approach for conceptual explicitation.
In *Model-Based Reasoning in Science and Technology – Inferential Models for Logic, Language, Cognition and Computation*, volume 49 of *SAPERE*, pages 441–469. Springer, 2019.

[3] A. Steen.
Extensional Paramodulation for Higher-Order Logic and its Effective Implementation Leo-III, volume 345 of *DISKI – Dissertations in Artificial Intelligence*.
Akademische Verlagsgesellschaft AKA GmbH, Berlin, 9 2018.

Semi-Automated Translation into Logical Formalism

In a semi-automated setting, the grammatical structures can be formalized into semantically equivalent entities in an expressive logical formalism. In the given example, the first tweet can be translated as:

$$\exists x.\,\mathrm{source}(x) \;\wedge\; \mathrm{credible}(x) \;\wedge\; \mathrm{calls}(x, \mathsf{officeOfTrump}) \\ \wedge\; \mathrm{SAYS}(x, \mathrm{isFraud}(\mathsf{birthCertOfObama}))$$

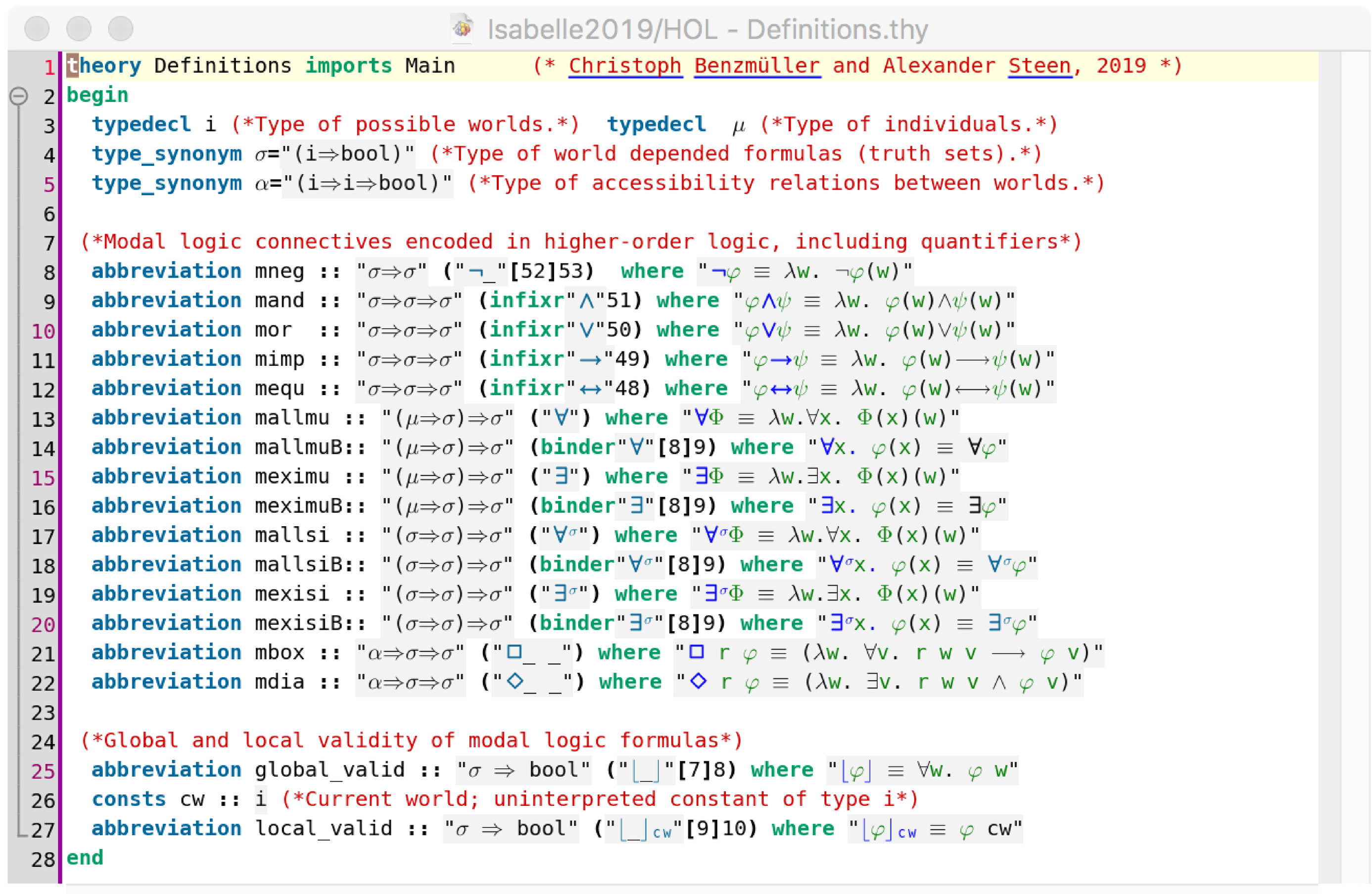
The first sentence of the second tweet is formalized as:

$$\mathrm{ANYTIME}\Big(\forall \mathit{story}.\,\Big((\mathrm{isAbout}(\mathit{story}, \mathsf{trump}) \vee \mathrm{isAbout}(\mathit{story}, \mathsf{campaignOfTrump})) \\ \wedge (\exists x.\,\mathrm{source}(x) \wedge \mathrm{SAYS}(x, \mathit{story}))\Big) \Rightarrow \forall y.\,\neg(\mathrm{BELIEF}(y, \mathit{story}))\Big)$$

The challenge is to adequately represent the notions of SAYS, ANYTIME, and BELIEF, which are no mere predicates but semantically complex modal operators with non-trivial properties. A quantified multi-modal logic is explored and utilised below (temporal, doxastic and indexical). Minimal assumptions about the combined modalities are needed so that the irrationality of the tweets can be revealed by automated theorem provers.

Argument Assessment via Automated Reasoning

Higher-order logic, when utilised as a meta-logic, is sufficiently expressive to encode and combine various object logics; properties of these object logics can be modified on the fly.



Isabelle2019/HOL - Definitions.thy

```
theory Definitions imports Main      (* Christoph Benzmüller and Alexander Steen, 2019 *)
begin
  typedecl i (*Type of possible worlds.*)  typedecl  μ (*Type of individuals.*)
  type_synonym σ="(i⇒bool)" (*Type of world depended formulas (truth sets).*)
  type_synonym α="(i⇒i⇒bool)" (*Type of accessibility relations between worlds.*)

 (*Modal logic connectives encoded in higher-order logic, including quantifiers*)
  abbreviation mneg :: "σ⇒σ" ("¬_"[52]53)  where "¬φ ≡ λw. ¬φ(w)"
  abbreviation mand :: "σ⇒σ⇒σ" (infixr"∧"51) where "φ∧ψ ≡ λw. φ(w)∧ψ(w)"
  abbreviation mor  :: "σ⇒σ⇒σ" (infixr"∨"50) where "φ∨ψ ≡ λw. φ(w)∨ψ(w)"
  abbreviation mimp :: "σ⇒σ⇒σ" (infixr"→"49) where "φ→ψ ≡ λw. φ(w)⟶ψ(w)"
  abbreviation mequ :: "σ⇒σ⇒σ" (infixr"↔"48) where "φ↔ψ ≡ λw. φ(w)⟷ψ(w)"
  abbreviation mallmu :: "(μ⇒σ)⇒σ" ("∀") where "∀Φ ≡ λw.∀x. Φ(x)(w)"
  abbreviation mallmuB:: "(μ⇒σ)⇒σ" (binder"∀"[8]9) where "∀x. φ(x) ≡ ∀φ"
  abbreviation meximu :: "(μ⇒σ)⇒σ" ("∃") where "∃Φ ≡ λw.∃x. Φ(x)(w)"
  abbreviation meximuB:: "(μ⇒σ)⇒σ" (binder"∃"[8]9) where "∃x. φ(x) ≡ ∃φ"
  abbreviation mallsi :: "(σ⇒σ)⇒σ" ("∀σ") where "∀σΦ ≡ λw.∀x. Φ(x)(w)"
  abbreviation mallsiB:: "(σ⇒σ)⇒σ" (binder"∀σ"[8]9) where "∀σx. φ(x) ≡ ∀σφ"
  abbreviation mexisi :: "(σ⇒σ)⇒σ" ("∃σ") where "∃σΦ ≡ λw.∃x. Φ(x)(w)"
  abbreviation mexisiB:: "(σ⇒σ)⇒σ" (binder"∃σ"[8]9) where "∃σx. φ(x) ≡ ∃σφ"
  abbreviation mbox :: "α⇒σ⇒σ" ("□_ _") where "□ r φ ≡ (λw. ∀v. r w v ⟶ φ v)"
  abbreviation mdia :: "α⇒σ⇒σ" ("◇_ _") where "◇ r φ ≡ (λw. ∃v. r w v ∧ φ v)"

 (*Global and local validity of modal logic formulas*)
  abbreviation global_valid :: "σ ⇒ bool" ("⌊_⌋"[7]8) where "⌊φ⌋ ≡ ∀w. φ w"
  consts cw :: i (*Current world; uninterpreted constant of type i*)
  abbreviation local_valid :: "σ ⇒ bool" ("⌊_⌋cw"[9]10) where "⌊φ⌋cw ≡ φ cw"
end
```

Reasoning Technology:

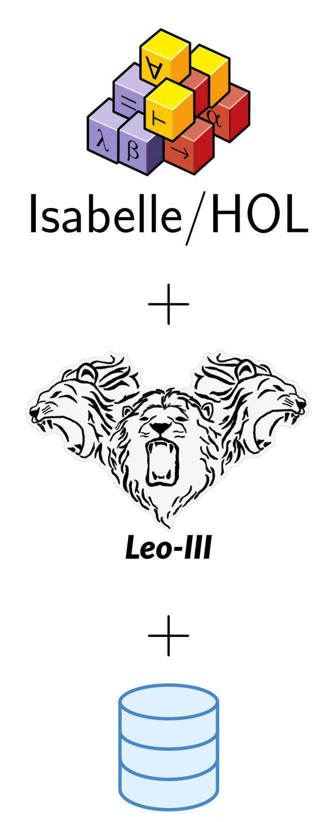


Figure 3: Encoding of quantified multi-modal logic in meta-logic HOL.

When assuming (a) that Trump believes what he says, and (b) that Trump's belief is reflecting what is the case, then a basic contradiction can be inferred.

Isabelle2019/HOL - TrumpInIsabelle.thy

```
theory TrumpInIsabelle                     (*Christoph Benzmüller and Alexander Steen, 2019*)
imports Definitions (*import of basic modal logic concepts in HOL *)
begin                      (*unimportant*) sledgehammer_params[type_enc=mono_native,verbose]

consts (*modal accessibility relation for time, belief (of an agent), speech (of an agent)*)
 AnyTimeRel::α   BeliefOfRel::"μ⇒α"   SaysRel::"μ⇒α"
abbreviation "ANYTIME φ ≡ □AnyTimeRel φ"        (*ANYTIME as modal operator*)
abbreviation "BELIEF x φ ≡ □(BeliefOfRel x) φ" (*BELIEF as modal operator*)
abbreviation "SAYS x φ ≡ □(SaysRel x) φ"        (*SAYS as modal operator*)
consts (*uninterpreted, scenario specific constant symbols*)
 Trump::μ campaignOfTrump::μ officeOfTrump::μ birthCertObama::μ calls::"μ⇒μ⇒σ" isAbout::"σ⇒μ⇒σ"
 source::"μ⇒σ" credible::"μ⇒σ" isFraud::"μ⇒σ"

axiomatization where  (*formalisation of Trump's tweets in multi-modal logic*)
 Tweet1: "⌊SAYS Trump (∃x. source x ∧ credible x ∧ calls x officeOfTrump ∧
                    SAYS x (isFraud birthCertObama))⌋" and
 Tweet2: "⌊SAYS Trump (ANYTIME
           (∀σs. ((isAbout s Trump ∨ isAbout s campaignOfTrump) ∧ (∃x. source x ∧ SAYS x s))
               → (∀y. ¬(BELIEF y s))))⌋" and
 Implicit: "⌊isAbout (isFraud birthCertObama) campaignOfTrump⌋" and (*implicit knowledge*)
 AnyTimeIncludesNow: "⌊∀σs. ANYTIME s → s⌋"  (*implicit knowledge*)

lemma True nitpick [satisfy,user_axioms,show_all] oops (*consistent? yes, model by nitpick*)

(*Trump (essentially) says that he does not belief that birthCertObama is fraud*)
lemma "⌊SAYS Trump (¬(BELIEF Trump (isFraud birthCertObama)))⌋"
  using Tweet1 Tweet2 AnyTimeIncludesNow Implicit sledgehammer() oops (*proof by ATPs*)

axiomatization where TrumpBelievesWhatHeSays: "⌊∀σs. SAYS Trump s → BELIEF Trump s⌋"  and
                     TrumpsBeliefIsTruth: "⌊∀σs. BELIEF Trump s ↔ s⌋"

(*Trump does not belief that birthCertObama is fraud (when assuming the latter axioms*)
lemma "⌊(¬(BELIEF Trump (isFraud birthCertObama)))⌋"
  using Tweet1 Tweet2 AnyTimeIncludesNow Implicit TrumpBelievesWhatHeSays TrumpsBeliefIsTruth
  sledgehammer[remote_leo3]() oops (*proof by ATPs*)
(*birthCertObama is no fraud*)
lemma "⌊(¬(isFraud birthCertObama))⌋"
  using Tweet1 Tweet2 AnyTimeIncludesNow Implicit TrumpBelievesWhatHeSays TrumpsBeliefIsTruth
  sledgehammer[remote_leo3]() oops (*proof by ATPs*)

axiomatization where
 TrumpBelievesCredibleSources: "⌊∀σs. (∃x. source x ∧ credible x ∧ SAYS x s) → BELIEF Trump s⌋"

lemma False  (*Trump is an inconsistent, irrational agent: Falsity follows*)
  using Tweet1 Tweet2 AnyTimeIncludesNow Implicit TrumpBelievesWhatHeSays TrumpsBeliefIsTruth
        TrumpBelievesCredibleSources sledgehammer[remote_leo3 e]() oops (*proof by ATPs*)
end
```

Figure 4: Formal assessment of Trump's tweets in Isabelle utilising universal meta-logical reasoning.

c.benzmueller@fu-berlin.de · alexander.steen@uni.lu