On the impediment of logical reasoning by non-logical inferential methods

Andreas Haida, Luka Crnič & Yosef Grodzinsky

ELSC & LLCC, The Hebrew University of Jerusalem

grodzinskylab.com

September 10, 2017

Sinn und Bedeutung 22

Special session *Semantics and Natural Logic*

Drawing inferences

- Assume that sentence (A) is true. Does it follow logically that (I) is true as well?

(A)	All of the Ms are Ks	✓ in 73% of all trials
(I)	Some of the Ms are Ks	

- Now, assume that sentence (I) is true. Can you logically infer that (O) is true, too?

(I)	Some of the Ms are Ks	✓ in 94% of all trials
(O)	Some of the Ms are not Ks	

- Finally, is the following inference logically valid?

(A)	All of the Ms are Ks	✗ in 98% of all trials
(O)	Some of the Ms are not Ks	

- The numbers show the results of Newstead and Griggs (1983).

Logical validity vs. perceived validity

This is how the observed behavior matches with logical behavior:

Inference	% accept expected in Aristotelian logic	% accept observed	% error
(A) to (I)	100	73	27
(I) to (O)	0	94	94
(A) to (O)	0	2	2

- What causes the two 'big' error rates?
  – Subjects compute scalar inferences (SIs).
- Why are the error rates not (close to) 100%?
  – There are different populations:
  Some reasoners compute SIs, some don't.
- Why different error rates (27 vs. 94%)?
  – Again, there are different populations:
  Some reasoners compute SIs for premises only.

SIs of existential sentences

- (I) sentences are systematically ambiguous:

(I)	Some of the Ms are Ks	
(I_w)	*There are Ms that are Ks*	(weak interpretation)
(I_s)	*Only some of the Ms are Ks*	(strong interpretation)

- (I_s) is derived from (I_w) by a SI:

  (I_s) $\equiv$ (I_w) $\wedge$ $\neg$*All of the Ms are Ks*

- The same holds for (O) sentences:

(O)	Some of the Ms are not Ks	
(O_w)	*There are Ms that are not Ks*	
(O_s)	*Only some of the Ms are not Ks*	$\equiv$ (I_s)

  (O_s) $\equiv$ (O_w) $\wedge$ $\neg$*All of the Ms are not Ks*

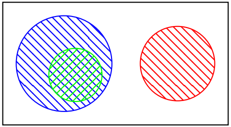
- To test our hypothesis that there are different groups of reasoners with respect to SI computation, we're investigating syllogistic reasoning.

Syllogisms

- Syllogisms are arguments of the form $\dfrac{\text{Premise}_1 \\ \text{Premise}_2}{\text{Conclusion}}$

- Here's an example of a valid syllogism:

(E)	No Italians are miners
(A)	All bikers are Italians
(E)	No bikers are miners



- New sentence type: (E) sentences; hence, 4 types overall
- Each syllogism has three terms in the premises. There are 4 possible arrangements of these three terms. The term arrangement in the conclusion is fixed:

- 4 sentence types for each of the 3 lines, and 4 arrangements: $4^3 \times 4 = 256$ syllogisms

How ambiguity impedes syllogistic reasoning performance

- Only 24 of the 256 syllogisms are valid in Aristotelian logic.
- Previous studies observed error rates of up to $> 80\%$ in performing syllogistic reasoning.
- These studies suggest that the linguistic ambiguity of existential sentences, i.e. of (I) and (O) sentences, impedes the reasoning performance.
- Importantly, the ambiguity of existential sentences can affect the (in)validity of a syllogism differently.
- This has been observed before (Rips 1994), but we're investigating this systematically.

Syllogism classes (granularity level I)

We can identify 6 classes, which we characterize in terms of how they are affected by SI computation.

There are 2 invariant classes:

- $[-v \overset{\text{SI}}{\rightsquigarrow} -v]$: Invalid syllogisms that are unaffected by SI computation
  - Invalid syllogisms without existential premises
  - Syllogisms that are invalid on all readings of their existential premises
- $[+v \overset{\text{SI}}{\rightsquigarrow} +v]$: Valid syllogisms that are unaffected by SI computation
  - Valid syllogisms with an (A) or (E) conclusion

Syllogism classes (granularity level I)

There are 4 variant classes:

- $[-v \overset{\text{SI}}{\rightsquigarrow} +v]$: Invalid syllogisms that are validated by SI computation
  - Invalid syllogisms with an existential premise (a necessary but not sufficient condition)
- $[+v \overset{\text{SI}}{\rightsquigarrow} -v]$: Valid syllogisms that are invalidated by SI computation
  - Valid syllogisms with an existential conclusion
- $[-v \overset{\text{SI}}{\rightsquigarrow} \pm v]$: Invalid syllogisms that are validated by the SI of a premise but only if the SI of the conclusion is not computed
  - Invalid syllogisms with an existential premise and an existential conclusion
- $[+v \overset{\text{SI}}{\rightsquigarrow} \pm v]$: Valid syllogisms that are invalidated by the SI of the conclusion but only if the SI of a premise is not computed
  - Valid syllogisms with an existential premise and an existential conclusion

An example: how to identify class $[-v \overset{\text{SI}}{\rightsquigarrow} +v]$

- (A) and (E) conclusions can be only be validated by (A) and (E) premises.
- Thus, class $[-v \overset{\text{SI}}{\rightsquigarrow} +v]$ can only contain syllogisms with (I) or (O) conclusions.
- However, the SI of the (I) or (O) conclusion must also be validated by the premises and their SIs, or else we end up in class $[-v \overset{\text{SI}}{\rightsquigarrow} \pm v]$.
- This means we need to find a pair of valid syllogisms that differ only in that one contains (I) sentences in places where the other contains (O) sentences.
- Luckily, Aristotelian logic gives us such a pair (but only one such pair): IA3I and OA3O.
- This means that class $[-v \overset{\text{SI}}{\rightsquigarrow} +v]$ has the following two members (and only these two members): IA3O and OA3I.
- Eventually, we wrote a theorem prover for Aristotelian logic to free us from such brain gymnastics.

Testing syllogistic reasoning performance

- We conducted an experiment with 120 participants over AMT.
- We restricted attention to 5 of the 6 classes.
- Each participant: 100 binary acceptability judgments for 20 tokens of each of 5 syllogism classes determined by the occurrence of existential sentences in premises and conclusion.
- Here are the mean acceptance rates of each class:

Class	% acc.
$[-v \rightsquigarrow -v]$	19.0
$[-v \rightsquigarrow \pm v]$	56.4
$[-v \rightsquigarrow +v]$	64.6
$[+v \rightsquigarrow -v]$	60.7
$[+v \rightsquigarrow +v]$	76.3

- Some of these results are easily interpretable: e.g., syllogisms in $[+v \overset{\text{SI}}{\rightsquigarrow} +v]$ are accepted more often than those in $[-v \overset{\text{SI}}{\rightsquigarrow} -v]$.
- But what to make of the observation that e.g. syllogisms in $[-v \overset{\text{SI}}{\rightsquigarrow} +v]$ are accepted more often than those in $[-v \overset{\text{SI}}{\rightsquigarrow} \pm v]$?

Hypothesis and prediction

- There here are three groups of reasoners:

	Logicians	Validators	Strength-eners	
Premise	weak (*There are ...*)	strong (*Only some ...*)	strong	validates an invalid argument
Conclu-sion	weak	weak	strong	invalidates a valid argument

- We expect to observe three different behavioral patterns:

Syllogism class	Logicians	Validators	Strengtheners	
$[-v \rightsquigarrow -v]$	✗	✗	✗	invariant
$[-v \rightsquigarrow \pm v]$	✗	✓	✗	affected by ambiguity
$[-v \rightsquigarrow +v]$	✗	✓	✓	
$[+v \rightsquigarrow -v]$	✓	✓	✗	
$[+v \rightsquigarrow \pm v]$	✓	✓	✗	
$[+v \rightsquigarrow +v]$	✓	✓	✓	invariant

Results

- Now let's take another look at the mean acceptance rates:

Class	L	V	S	% acc.
$[-v \rightsquigarrow -v]$	✗	✗	✗	19.0
$[-v \rightsquigarrow \pm v]$	✗	✓	✗	56.4
$[-v \rightsquigarrow +v]$	✗	✓	✓	64.6
$[+v \rightsquigarrow -v]$	✓	✓	✗	60.7
$[+v \rightsquigarrow +v]$	✓	✓	✓	76.3

- Green links highlight the observations that we correctly predict. E.g., the acceptance rate of $[-v \overset{\text{SI}}{\rightsquigarrow} +v]$ is higher than that of $[-v \overset{\text{SI}}{\rightsquigarrow} \pm v]$ because of the population of strengtheners.
- However, there's also a red link, where we fail: Because of the logicians, we expect syllogisms in $[-v \overset{\text{SI}}{\rightsquigarrow} \pm v]$ to be accepted less often than syllogisms in $[+v \overset{\text{SI}}{\rightsquigarrow} -v]$; however, the difference doesn't reach significance.

Towards a more fine-grained classification

No significant difference between $[-v \overset{\text{SI}}{\rightsquigarrow} \pm v]$ and $[+v \overset{\text{SI}}{\rightsquigarrow} -v]$

- Reason: there's a lot of variation accross the syllogisms in $[+v \overset{\text{SI}}{\rightsquigarrow} -v]$.
  For example:
  - AI3I, IA4I: accepted ~ 80% of all times
  - AE4O, EA2O: only accepted ~ 50% of all times
- Where does this variation come from?
  - AI3I, IA4I: the SI of the conclusion invalidates the syllogism.
  - AE4O, EA2O: the SI is <u>inconsistent</u> with the premises.
- Hypothesis: Inconsistency leads to better recognition of invalidity.
- To test this hypothesis, our classification needs to take inconsistency into account.

A more fine-grained classification

Taking inconsistency into account leads to the following subclassifications:

- Subclass $[+v \overset{\text{SI}}{\rightsquigarrow} -c]$ of $[+v \overset{\text{SI}}{\rightsquigarrow} -v]$

Class	% acc. our data	% acc. Rips (1994)
$[+v \rightsquigarrow -v]$	71.3%	65%
$[+v \rightsquigarrow -c]$	51.9%	57%

- Subclass $[-v \overset{\text{SI}}{\rightsquigarrow} -c]$ of $[-v \overset{\text{SI}}{\rightsquigarrow} -v]$
- Subclass $[-c]$ of $[-v \overset{\text{SI}}{\rightsquigarrow} -v]$: Syllogisms that are formed from sets of inconsistent sentences (counterparts of valid syllogisms, where the valid conclusion is replaced by the contradictory sentence).

Class	% acc. Rips (1994)
$[-v \rightsquigarrow -v]$	10.3%
$[-v \rightsquigarrow -c]$	1.5%
$[-c]$	1%

Do the means reflect subpopulations?

- We expect that taking (SI induced) inconsistency into account will give us all the predicted differences between means.
- Let's assume that we'll indeed get the following result:

Class	L	V	S	% acc.
$[-v \rightsquigarrow -v]$	✗	✗	✗	m_1
$[-v \rightsquigarrow \pm v]$	✗	✓	✗	m_2
$[-v \rightsquigarrow +v]$	✗	✓	✓	m_3
$[+v \rightsquigarrow -v]$	✓	✓	✗	m_4
$[+v \rightsquigarrow +v]$	✓	✓	✓	m_5

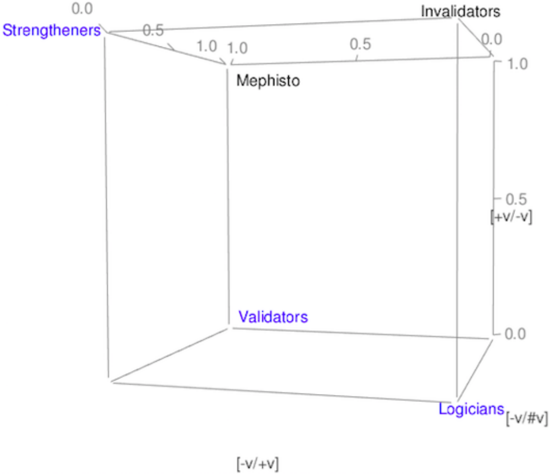
- How can we show that the means reflect homogeneous behavior within different groups and not heterogeneous behaviour of all subjects?
- Recall that every subject judged (will judge) 20 instances of each of the 5 syllogisms classes.
- This means that for every subject we have a rich behavioral profile so that we can detect (in)consistent behavior.

Do the means reflect subpopulations?

- To identify subpopulations, we used a density-based clustering algorithm: DBSCAN.
- We'll show you what DBSCAN gives us for the data of our AMT experiment.
- One thing you'll see is that the data is very noisy.
- In the AMT experiment, the reaction times show that most subjects started to give very quick responses after a while.
- To prevent this, we'll conduct our next experiment in the lab.
- However, even through the noise we can see that one of our hypothesized groups seems not to exist.

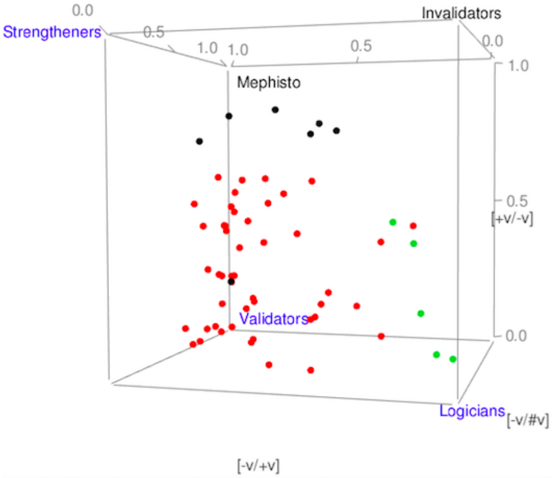
Identifying groups of reasoners

- The behavior towards the 2 invariant classes gives a measure of a subject's logical abilities.
- The behavior towards the 3 variant classes is represented by the deviance from the subject's logical abilities.



The results: two populations

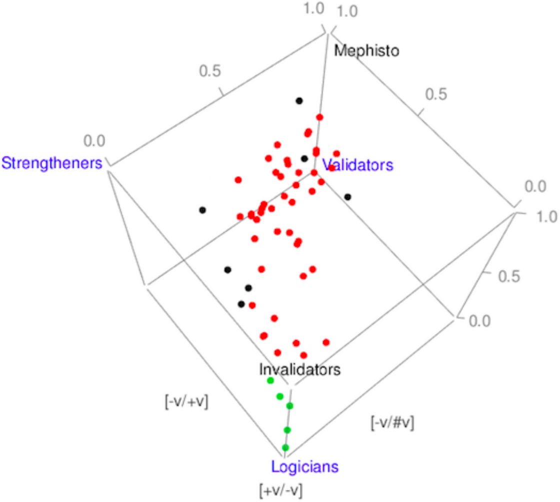
- We eliminated subjects with $> 12.5\%$ error rate in the invariant classes (half of all subjects).
- Two density clusters: red and green; outliers are black
- Members of the green cluster are in the neighborhood of the logicians' corner.



- A large group of members of the red cluster is in the neighborhood of the validators' corner.
- The strengtheners' corner is not populated.

The results: no systematic strengthening of conclusions

- Left of the diagonal: subjects that strengthen conclusions sometimes
- But: no systematic strengthening of conclusions; i.e. no strengtheners
- No evidence for other populations



Conclusions

What we did:

(i) We developed a quantitative method to study tendencies across syllogism types.

(ii) We showed evidence for the existence of groups of reasoners.

(iii) We identified a supra-sentential context in which some subjects systematically do <u>not</u> compute SIs.

Overall, we observe behavior which is grounded in logical reasoning and natural language interpretation.

We found initial evidence for two groups of reasoners:

- subjects who consistently employ Aristotelian logic and don't compute SIs (logicians)
- subjects who consistently employ Aristotelian logic and maximize derivable inferences by computing SIs for premises but not for conclusions (validators).

Appendix: How to explain the behavior of validators?

In 27% of all (A) to (I) trials, (I) is interpreted as (I_s).

In 94% of all (I) to (O) trials, (I) is interpreted as (I_s).

- We saw evidence that there are validators.
- Let's assume that the difference above is due to this group of reasoners: they don't compute the SI of the (I) conclusion.
- How can we explain this behavior?
- We'll discuss a linguistically interesting hypothesis, and why it cannot be maintained for syllogistic reasoning.

Appendix: A hypothesis regarding the behavior of validators

- SIs serve to eliminate speaker ignorance inferences (Fox 2007).
- Validators take the premise(s) and conclusion of an argument as utterances of one and the same speaker.
- Here is how the first assumption leads to (I) to (O) inferences:

(I_w) Some Ms are Ks

'All Ms are Ks' is a relevant alternative and not settled by (I_w)

$\rightsquigarrow$ The speaker is ignorant about 'All Ms are Ks' (by quantity)

(SI) ¬All Ms are Ks (from (I_w) to eliminate the ignorance inf.)

(O_s) Some Ms are not Ks (by (I_w) and (SI))

- Together, the two assumptions inhibit (A) to (I) inferences:

(A) All Ms are Ks

(I_w) Some Ms are Ks (by the Aristotelian meaning of (A))

'All Ms are Ks' is a relevant alternative of (I_w)

'All Ms are Ks' is entailed, hence settled by (A)

No ignorance inference from (I_w) by quantity reasoning

No motivation to compute a SI for (I_w)

Appendix: A hypothesis regarding the behavior of validators

- Unfortunately, this account is not supported by the syllogism data that we have:

Class	% acc. our data	% acc. Rips (1994)
$[+v \rightsquigarrow -v]$	71.3%	65%
$[+v \rightsquigarrow -c]$	51.9%	57%

- Being a member of $[+v \rightsquigarrow -c]$ means that the premises entail the contradiction of the SI of the conclusion.
- Thus, the premises settle the stronger alternative to the conclusion.
- Given our assumptions, this means that there is no motivation for validators to compute the SI in the first place.
- Thus, our assumptions lead us to expect that syllogisms in $[+v \rightsquigarrow -c]$ are accepted more often than syllogisms in $[+v \rightsquigarrow -v]$, contrary to fact.