

Syllogistic reasoning as communication

Michael Henry Tessler, Noah D. Goodman

manuscript draft

Reasoning draws the conclusion and makes us grant the conclusion, but *does not make the conclusion certain*. Nor does it remove doubt so that the mind may rest on the intuition of truth, unless the mind discovers it by the path of experience.

Roger Bacon, *Opus Majus* (1267)

The syllogism is like a text: fixed, boxed-off, isolated... The riddle [by contrast] belongs in the oral world. To solve a riddle, canniness is needed: one draws on knowledge, often deeply subconscious, beyond the words themselves in the riddle.

Walter J. Ong, *Orality and Literacy* (1982)

Imagine you're discussing the 2014 World Cup with your friend. Your friend notices and remarks:

"*All* of the teams that made it to the semifinals were expected to advance; they had favorable standings before the tournament. And did you notice that *some* of the teams expected to advance had favorable referees? Remember the opening match: Brazil versus Croatia?"

At this point, you might be willing to conclude that probably some of the teams that made it to semifinals had favorable referees. The argument seems reasonable, but it isn't valid in the deductive sense. Why does it seem reasonable?

1 Introduction

Your friend's argument sounds like a syllogism: a two sentence argument that uses quantifiers as relations (e.g. *all*, *some*) between terms (e.g. *semifinalists*, *teams expected to advance*). We can represent the syllogism without the content (Table 1, left) and without the relations between terms (i.e. just as a logical form; Table 1, right).

All A are B	A – B
Some B are C	B – C
—	—
Some A are C	A – C

Table 1: World cup syllogism, without content [left] and without quantifiers [right]

The syllogism is considered the first formal system of reasoning, developed by Aristotle in the 4th century BC. Syllogistic reasoning has been replaced by more modern formalisms (e.g. the predicate calculus), yet persists in the experiments of cognitive psychologists. Syllogisms seem like an intriguing testing ground for human reasoning because they (a) are a formal system of logic and (b) use natural language in their construction. The former ensures there is a normatively correct solution, while the latter allows people with no training in formal logic to analyze these arguments.

2 The variety of syllogistic reasoning tasks

3 The argument strength of a syllogism

We begin with the assumption that reasoning with language, of which syllogisms are a special case, is no different than reasoning in everyday contexts. We follow in a tradition suggesting that everyday reasoning is uncertain and best described by the tools of probability theory. In this line of thinking, a deductive argument is understood as an argument that is maximally strong, with a spectrum of argument strength existing below it, given by $\Pr(\text{conclusion} \mid \text{premises})$ (Oaksford & Chater, 2007; Lassiter & Goodman, 2014).

3.1 Not all invalid syllogisms are weak arguments

Evans and colleagues (1999) had participants rate all possible combinations of 64 syllogistic premises and 4 conclusions using an evaluation task with abstract content (letters e.g. D, H, Z for terms). The study was aimed at examining the differences between instructions (what is a *{necessary, possible}* conclusion) to test some predictions of the mental models framework.

The investigators discovered unexpectedly that among syllogisms with possible (but not necessary) conclusions (i.e. *fallacies*, in the deductive sense), some are very strongly endorsed (in both instructional conditions) while others are very weakly endorsed. They replicated this finding in a separate experiment, in which participants were presented with four problem types: Necessary (i.e.

valid), Possible Strong (PS), Possible Weak (PW), and Impossible (a contradiction of a Necessary conclusion). They found that the PS problems were endorsed about as often as the Necessary problems, and the PW problems were endorsed almost as little as Impossible problems.

The notion of Possible Strong and Possible Weak syllogisms falls right out of analyzing the argument strength of syllogisms.

[Plot here a histogram/density-plot of argument strength of 256 syls X 2 term orderings and highlight the possibly weak and possible strong ones (i.e. from the Evans study).]

4 The influence of background knowledge

The contents of a syllogism affect syllogistic reasoning (Wilkins, 1928). This has traditionally been explored in the interaction between logical validity and the *a priori* believability of the conclusion. The effect is most prominent in the syllogistic evaluation task and has loosely been described as a tendency to endorse *a priori* believable conclusions, regardless of the logical validity of the argument. However, the degree to which conclusion-believability influences acceptance rates is more marked on invalid than on valid syllogisms (J. S. B. T. Evans, Handley, & Pollard, 1983). When experimenters have included neutral materials for baseline comparisons, they find belief bias is primarily associated with *rejecting* unbelievable conclusions particularly when the argument is fallacious (i.e. an increase in correct rejections for invalids), leading some investigators to refer to it as “belief debias” (Morley, Evans, & Handley, 2004; Newstead, Pollard, Evans, & Allen, 1992).

Evans and colleagues (2001) followed up on the J. S. B. T. Evans et al. (1999) study that found different endorsement rates for Possible Weak and Possible Strong syllogisms. The investigators hypothesized that PW problems would exhibit a positive belief bias (enhancement of endorsements for believable, invalid conclusions — relative to neutral conclusions) and PS problems would exhibit a negative belief bias (the typical “belief debias”). This predictions follow from the J. S. B. T. Evans et al. (1999) finding that with neutral or abstract content, PS endorsements are near ceiling and PW endorsements are near floor.

4.1 Extant theories

Much of the theoretical discussion on belief bias is concerned with *the stage of processing* where the influence of background knowledge is taken into account. According to this, theoretical accounts can be understood as arguing that background knowledge is incorporated during (1) the encoding of the problem (the translation of the syllogism into a mental representation); (2) the reasoning process (the manipulation of the mental representation); or (3) the decoding of the reasoning process (the translation of the mental representation into a response).

4.1.1 Some old qualitative theories

Selective scrutiny, misinterpreted necessity, mental models

4.1.2 Recent quantitative theories

Two recent investigations of belief bias have used computational models to disambiguate a stage 2 effect from a stage 3 effect.

Klauer, Musch, and Naumer (2000) used a Multinomial Process Tree (MPT) model to argue that the belief bias effects occur during the reasoning process (stage 2). This comes directly from the model itself, which models the task as consisting of either (i) accurately reasoning, i.e. decided whether or not the problem is valid or invalid, and responding correctly (with probability determined by model parameter $r_{problem.type}$) or (ii) guessing (with probability $1 - r_{problem.type}$). Guessing leads to a guessing subtree, where it is assumed prior beliefs about the conclusion can have influence. Prior beliefs are modeled using parameters $\beta_{believable}$ and $\beta_{unbelievable}$. Problem type (for the reasoning parameters) consists of four possibilities, resulting from crossing believability and validity; hence, there are 4 r parameters, and these are used to test the hypothesis that the *reasoning process* is different across the 4 conditions. There are two β parameters corresponding to believable and unbelievable response biases, and these are used to test if beliefs influence the response stage. The investigators found significant differences between the reasoning parameters across conditions, while constraining the β belief parameters to be equal had little effect on the fit. From this, the authors concluded belief bias is a largely an effect on the reasoning process, consistent with dual-process theories of reasoning.

Dube, Rotello, and Heit (2010) drew issue with the use of Klauer et al.'s MPT models because the models assumed a "simple threshold". Dube et al. pointed out that the assumptions of threshold models are the same as the assumptions of measuring accuracy by *Hits - False Alarms*: that, for a given problem type and a constant level of accuracy, changes in response bias are associated with *equal changes* in acceptance rates. If the relationship between response bias and acceptance rates is empirically nonlinear however, the above assumption could lead one to infer a difference in sensitivity, when only a difference in bias is present (i.e. a Type I error on the null hypothesis that there is no difference in sensitivity / accuracy between belief conditions). To interrogate this assumption, the investigators measured ROC curves by using confidence ratings following a "valid"/"invalid" judgment. They found that the assumption of linear ROCs is unwarranted, calling into question all current models of belief bias (which implicitly use this assumption).

The authors went on to argue that the "belief bias" effect is a "response bias" effect. They do this using two null results. First, they observe that the points for believable and unbelievable ROC curves appear to lie on a single curve, "indicating subjects showed little to no difference in accuracy when

judging conclusion validity”. They compare the estimated area under the ROC curve for the two conditions using A_z , a SDT statistic used for accuracy when the equal-variance assumption cannot be applied¹. They find no significant difference in the estimated area under the two ROCs. Second, they constrain the “reasoning” parameters of their Signal Detection Model ($d_{believable}$ and $d_{unbelievable}$; the distance between valid and invalid distributions of argument strength in the 2 belief conditions) to be equal. They observe a non-significant effect on the fit of the SDT models, “indicating a negligible effect of believability on accuracy”.

4.2 Belief and argument strength

For both of these approaches, it’s important to bear in mind the particular materials used to elicit the effect. As we’ve suggested above, the syllogistic space actually defines a distribution over argument strengths. Qualitative differences present in this distribution have been talked about before in terms of “single model” vs. “multiple model” problems (a distinction among valid syllogisms) (Johnson-Laird & Byrne, 1991) as well as “strong possible” and “weak possible” (a distinction among invalid syllogisms) (J. S. B. T. Evans et al., 1999).

The argument strength distribution depends critically on the prior distribution over situations. Though the argument strength of a valid syllogism is always 1, the argument strength of an invalid syllogism will vary with the content. Consider the content used in the recents studies of belief bias by Dube et al. and Klauer et al. The syllogistic content was chosen so that the subject would not have strong prior beliefs about the truth of the premises (e.g. by choosing an esoteric or nonsense middle term, *some birds are metazoans; no sparrows are metazoans*), while having strong beliefs about the truth of the conclusion (*some birds are not sparrows*). We can import such a prior distribution over properties – $\Pr(bird, sparrow, metazoan)$ – into the model of argument strength to examine how such a distribution over properties shapes the prior distributions over sentences as well as the posterior distribution over argument strengths.

From the probabilistic perspective, the prior probability of *some birds are not sparrows* is very high. It’s hard to know exactly what you could say to a person to convince them that *some birds are not sparrows* is not the case. This is reflected in the number of invalid syllogisms for which *some birds are not sparrows* is still highly probable [data needed].

The same argument applies for *a priori* false (or, low probability) conclusions (e.g. *some sparrows are not birds*). This is reflected in the distribution of argument strength for this content over all 64 syllogisms [plot needed]. The distribution is heavily skewed towards the end-points. The prior probability of the conclusion is so low (or so high) that there is almost nothing you could tell a person to convince them otherwise.

¹Dube et al. mention A_z is an unbiased estimator of proportion correct in a 2AFC, and that it has smaller standard error than d_a in simulations by Mamillian et al. (2004)

4.2.1 Implications for previous studies

All of the invalid syllogisms used in Dube et al. (2010) would be classified as “strong possible” problems, with endorsement rates around 80% for plausibility judgments and 50% for necessity judgments using abstract content (J. S. B. T. Evans et al., 1999). J. S. Evans et al. (2001) showed that “strong possible” problems were most susceptible to “debiasing” effects, wherein the argument is correctly rejected when an unbelievable conclusion is used.

To put this in quantitative terms, we can analyze the informational content of the syllogism, as defined by the degree to which the syllogism updates our prior beliefs over sentences (into our posterior beliefs — the argument strength distribution). We can formalize this by using the expected KL divergence between a model of argument strength and a model of prior beliefs over sentences. We see that given the priors elicited from the content of previous belief bias experiments (Klauer, Dube), there is [probably, what we’ll see is a bimodal distribution, with high information content for valid syllogisms improbable conclusions, and low informational content for everything else]. The invalid syllogisms used by Dube et al. (2010) all have very low informational content. These are relatively poor experiments for distinguishing a model of reasoning (the argument strength model) from a model of not-reasoning (the prior distribution over sentences). Thus, any interpretations about the relative contribution of a “reasoning stage” based on these syllogisms should be called into question.

4.3 The probabilistic approach / the way forward

Studies of belief bias in syllogistic reasoning have looked at categorical distinctions between “believable” and “unbelievable” statements. In the most extreme cases, “unbelievable” conclusions are taken to be those which are literally false of the real world.

We began our discussion of belief bias with the statement that the content of a syllogism can affect the conclusions drawn. In a general sense, if we consider the syllogistic sentence — a statement about two terms or properties — the properties can be considered independent (e.g. novel words: *blickets* and *tomas*) or there can be some correlation between the terms (e.g. *religious people* and *church-goers*). Quantitatively, correlations can vary between -1 and +1. So far, most (possibly all [?]) studies of belief bias in syllogistic reasoning explore only the extreme end points of this spectrum.

Given the results of the argument-strength models above, it would be unwise to explore potentially intricate interactions between logic and belief using content that is either necessary (*some birds are not sparrows*) or impossible (*some sparrows are not birds*) given prior knowledge. This is because for [most / many / all?] syllogisms, the argument strength either implies one must disregard the content (valid syllogisms) or be driven by your prior beliefs (invalid syllogisms). It would be

more useful to see how prior beliefs can shape reasoning.

4.4 Experiment 2

We set out to explore how the content of the syllogistic argument can affect conclusions drawn. For this, we want use domains over which people have background knowledge while staying away from the endpoints of the *believable* – *unbelievable* spectrum.

4.4.1 Method

4.4.1.1 Prior elicitation

Causal knowledge is ideal for studying belief in reasoning because (1) people have strong intuitions about causal domains and (2) causal knowledge is uncertain, thus keeping us away from the extreme endpoints of the *believable* – *unbelievable* spectrum [citations?].

We began with causal domains that fell into two structural forms: common cause and multiple cause. For each structure, we used **some number** of different domains.

We based our selection of domains on simulations of argument-strength using qualitatively different priors (elicited from people in the lab). Our simulations suggested that domains using: **{common-cause / multiple-cause}** with **{2-enabling / 2-preventative / 1-enabling, 1-preventative}**.

We asked participants on Mechanical Turk to rate the likelihood of the properties co-occurring. We had participants rate all 8 combinations of properties.

4.4.1.2 Syllogism selection

We compared the posterior distribution of argument-strength using the empirically measured priors with the distribution of argument-strength using the best-fit `base-rate` parameter as an i.i.d. variable that generated situations (i.e. a model without structured background knowledge... a *naive* model) and the distribution on true conclusion sentences using the empirically measured priors (i.e. a model of just beliefs, no reasoning).

We examined the model predictions using conclusion terms coming from different parts of the structural form (e.g. in a common cause model, we considered arguments where the conclusion is between a cause and an effect as well as between the two effects).

We examined the expected information gain for each syllogism with the goal of disambiguating the 3 models (belief w/ reasoning, belief w/o reasoning, reasoning w/o beliefs).

4.4.1.3 Participants

4.4.1.4 Materials

4.4.1.5 Instructions

5 Formal semantics and generalized quantifiers

6 Pragmatics in syllogistic reasoning

6.1 Can nothing ever actually follow?

References

- Dube, C., Rotello, C. M., & Heit, E. (2010, July). Assessing the belief bias effect with ROCs: it's a response bias effect. *Psychological review*, 117(3), 831–63. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/20658855> doi: 10.1037/a0019634
- Evans, J. S., Handley, S. J., & Harper, C. N. (2001, August). Necessity, possibility and belief: a study of syllogistic reasoning. *The Quarterly journal of experimental psychology. A, Human experimental psychology*, 54(3), 935–58. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11548042> doi: 10.1080/713755983
- Evans, J. S. B. T., Handley, S. J., Harper, C. N. J., & Johnson-laird, P. N. (1999). Reasoning About Necessity and Possibility : A Test of the Mental Model Theory of Deduction. *Journal of experimental psychology. Learning, memory, and cognition*, 25(6), 1495–1513.
- Evans, J. S. B. T., Handley, S. J., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition*, 11, 295–306.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction* (L. Erlbaum, Ed.). UK: Hove.
- Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, 107(4), 852–884. Retrieved from <http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.107.4.852> doi: 10.1037/0033-295X.107.4.852
- Lassiter, D., & Goodman, N. D. (2014). How many kinds of reasoning ? Inference , probability , and natural language semantics. *submitted*(2014), 1–22.
- Morley, N. J., Evans, J. S. B. T., & Handley, S. J. (2004, May). Belief bias and figural bias in syllogistic reasoning. *The Quarterly journal of experimental psychology. A, Human experimental psychology*, 57(4), 666–92. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15204128> doi: 10.1080/02724980343000440

- Newstead, S. E., Pollard, P., Evans, J. S., & Allen, J. L. (1992, December). The source of belief bias effects in syllogistic reasoning. *Cognition*, 45(3), 257–84. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1490324>
- Oaksford, M., & Chater, N. (2007). *Bayesian rationality: the probabilistic approach to human reasoning*. Oxford: Oxford University Press.
- Wilkins, M. C. (1928). *The Effect of Changed Material on Ability to do Formal Syllogistic Reasoning*.