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

Andreas Haida, Luka Crnič & Yosef Grodzinsky
The Hebrew University of Jerusalem

Syllogistic reasoning is an element of Aristotle's philosophy of science. It serves to distinguish logically valid (+*v*) from logically invalid (–*v*) arguments. See (1) and (2) for examples:

(1) All desires are sins, some beliefs are desires; therefore, some beliefs are sins. (+*v*)

(2) All desires are sins, no desires are beliefs; therefore, some beliefs are not sins. (–*v*)

There is wealth of experimental studies investigating the performance in syllogistic reasoning tasks. These studies show mean error rates of up to 80% and more in judging the (in)validity of syllogistic arguments such as (1) and (2) (e.g. Dickstein 1978). This has sometimes been taken to show that humans do not engage in logical reasoning at all (e.g. by Chater and Oaksford 1999). Other researchers concluded that people do perform logic-based reasoning but also employ inferential methods that support logically invalid inferences or obviate logically valid inferences (e.g. Chapman and Chapman 1959; Dickstein 1978; Rips 1994; Geurts 2003). Two such methods are scalar-inference (SI) computation and illicit conversion (IC):

(3) a. some X are Y $\overset{\text{SI}}{\rightsquigarrow}$ not all X are Y b. some X are not Y $\overset{\text{SI}}{\rightsquigarrow}$ not all X are not Y

(4) a. all X are Y $\overset{\text{IC}}{\rightsquigarrow}$ all Y are X b. some X are not Y $\overset{\text{IC}}{\rightsquigarrow}$ some Y are not X

Note that (1) is invalidated by the SI of the conclusion (i.e., the two premises do not entail the SI of the conclusion); (2) is validated by the IC inference of the first premise (i.e., the IC inference of the first premise, in conjunction with the second premise, entails the conclusion). The study to be presented here is the first to analyze the use of these two inferential methods in individual performance. Our study reveals a systematic relationship between the use of the two methods and individual variation therein, the existence of two groups of reasoners, an effort-related response bias, and the preferred loci for SI computation in a supra-sentential context.

Method: To measure the impact of SI computation, we identified five classes in the classical set of syllogistic arguments (256 form types of which 24 are valid in Aristotelian logic). Specifically, we identified two classes that are invariant to SI computation, viz. (i) invariably invalid and (ii) invariably valid syllogisms, and three classes that are affected by SI computation, viz. (iii) invalid syllogisms that are validated by the SI of one of their premises, (iv) valid syllogisms that are invalidated by the SI of their conclusion, and (v) invalid syllogisms that are validated by the SI of a premise but only if the SI of the conclusion is not computed (i.e., the syllogisms of class (iii) are rendered +*v* by the SI of a premise whether or not the SI of the conclusion is computed). To quantify the impact of IC, half of the items in class (i) were chosen to be invariant to IC, subclass (i-a), and the other half to be validated by IC, subclass (i-b). Similarly, a proportion of the items in class (iv) were so chosen that IC performed on a premise reverses the invalidation of SI computation performed on the conclusion, defining subclass (iv-a) and (iv-b). (For balancing reasons, the other classes were not divided in this way.) This classification is shown by the table above. The last column gives a representative type in standard notation.

Class	Characteristic	Ex.
(i-a)	$[-v \overset{\text{SI}}{\rightsquigarrow} -v \overset{\text{IC}}{\rightsquigarrow} -v]$	IO4A
(i-b)	$[-v \overset{\text{SI}}{\rightsquigarrow} -v \overset{\text{IC}}{\rightsquigarrow} +v]$	AA2A
(ii)	$[+v \overset{\text{SI}}{\rightsquigarrow} +v]$	AA1A
(iii)	$[-v \overset{\text{SI}}{\rightsquigarrow} +v]$	IA3O
(iv-a)	$[+v \overset{\text{SI}}{\rightsquigarrow} -v \overset{\text{IC}}{\rightsquigarrow} -v]$	EA2O
(iv-b)	$[+v \overset{\text{SI}}{\rightsquigarrow} -v \overset{\text{IC}}{\rightsquigarrow} +v]$	IA4I
(v)	$[-v \overset{\text{SI}}{\rightsquigarrow} \pm v]$	AO1I

Regarding SI computation, we put forward the hypothesis that there are three groups of reasoners: those that do not compute any SIs (Logicians), those that compute SIs for premises but not for conclusions (Validators), and those that compute SIs for both premises and conclusions (Strengtheners). The table on page 2 shows the behavioral patterns exhibited by idealized members of these groups (where ✓ stands for accept as valid, ✗ for reject as valid, and ✗/✓ for accept as valid iff IC is performed on one of the premises). Furthermore, we expected to observe an effect of the use of IC in class (i-b) and (iv-b), driven by a hypothesized proportion of subjects that perform IC on premises, and no effect in the other classes (since they are invariant to IC). We conducted a forced-choice experiment in which participants had to accept/reject as valid 20 syllogism tokens of each of class (i) - (v) (100 judgments × 119 participants on Amazon MTurk). The mean acceptance rate of each class is shown in the table below. The last column displays the correlation between a subject's IC usage rate and the subject's acceptance rate of syllogisms of the respective class. (A subject's IC usage rate is given by the difference between the number of syllogisms the subject accepts in (i-b) and (i-a). We call the corresponding variable %IC.)

Class	L	V	S	% acc.	r
(i-a)	✗	✗	✗	19.0	—
(i-b)	✗/✓	✗/✓	✗/✓	52.8	—
(ii)	✓	✓	✓	76.3	.51
(iii)	✗	✓	✓	64.6	.43
(iv-a)	✓	✓	✗	55.3	–.1
(iv-b)	✓	✓	✗/✓	77.9	.48
(v)	✗	✓	✗	56.4	.28

Results: Since the judgments with respect to class (i-a) and (ii) are not impeded by SI computation or IC, the error rates of 19% false positives and 23.7% false negatives are the most immediate reflexes of real performance errors. They show that there is no general positive response bias. The green cells in the table show which observations are consistent with our hypotheses. (With respect to the acceptance rates, the ANOVA and post-hoc tests show that all and only those differences are significant that are greater than 2.5 pp.) The acceptance rate in the red cell runs counter to our expectation that syllogisms in (iv-a) are accepted more often than syllogisms in (v) (because of the hypothesized group of logicians). For class (ii), (iii), and (v), the observed correlation between %IC and individual acceptance rates is not expected.

Discussion: We propose that the unexpected correlations are explained by a positive (or anti-negative) answer bias that is triggered when a subject makes an effort to validate (by performing IC on a premise) and perceives a success (by recognizing a syllogism as valid regardless if the effort was in fact necessary); when no such effort is made, positive responses are sometimes supressed. The lack of a difference between the acceptance rates of (iv-a) and (v) can also be explained by this bias: In the case of (iv-a), r = –.1 suggests that more than half of all times that a subject performs IC on a premise the subject also computes the SI of the conclusion. Hence, less than half of all efforts are accompanied by the perception of success. In the case of (v), a higher proportion of efforts is accompanied by success (as suggested by r = .28). Thus, the positive answer bias lifts the acceptance rate of class (v) more than that of class (iv-a). This explanation is corroborated by the fact that elimination of subjects with %IC > 30% eliminates almost all subjects that accept more syllogisms of class (v) than of class (iv-a).

Identification of groups: For each subject x, the difference Δ_x between the number of syllogisms x accepts in (ii) and (i-a) is a measure of x's logical abilities. With Δ_x, x's performance relative to the other classes is gauged by the degree of deviance from Δ_x. This allows us to to identify behavioral patterns by abstracting from individual differences in logical abilities. Our identification method confirms the hypothesis of three groups of reasoners only partially: there is a small group of subjects (≈ 5% of all subjects) that qualify as logicians, a larger group of subjects (10% – 15%) that qualify as validators (both determined by a density-based clustering algorithm, DBSCAN), but no subjects that qualify as stregtheners. That is, we do find strengthening behavior (as shown by the difference in acceptance rate between class (ii) and (iv-a)) but subjects compute SIs consistently only for premises.

On the existence of a natural logic: In response to the existence question in the call for papers we give the following answer: In logical reasoning performance, we observe two consistent behaviors (that of logicians and validators). Both behaviors can be characterized by systems of inferential methods. Hence, there are two systems that might be called natural logic. One natural logic is Aristotelian logic (or first-order logic with non-empty domains of quantification). The other – more common – natural logic can be characterized by how it deviates from Aristotelian logic, viz. by the drive to increase derivable inferences through semantic/pragmatic enrichment of premises.

References

Chapman, Loren J., and Jean P. Chapman. 1959. Atmosphere effect revisited. Journal of Experimental Psychology 58:220–226.

Chater, Nick, and Mike Oaksford. 1999. The probability heuristics model of syllogistic reasoning. Cognitive Psychology 38:191–258.

Dickstein, Louis S. 1978. Error processes in syllogistic reasoning. Memory and Cognition 6:537–543.

Geurts, Bart. 2003. Reasoning with quantifiers. Cognition 86:223–251.

Newstead, Stephen E., and Richard A. Griggs. 1983. Drawing inferences from quantified statements: A study of the square of oppositions. Journal of Verbal Learning and Verbal Behavior 22:535–546.

Rips, Lance J. 1994. The psychology of proof. Cambridge, Mass.: MIT Press.