

Content Effects in Conditional Reasoning: Evaluating the Container Schema

Amber N. Bloomfield (a-bloomfield@northwestern.edu)

Department of Psychology, 2029 Sheridan Road
Evanston, IL 60208 USA

Lance J. Rips (rips@northwestern.edu)

Department of Psychology, 2029 Sheridan Road
Evanston, IL 60208 USA

Abstract

This study presented participants with conditional statements with content that was group-related, container-related, and a combination of group and container. Participants received modus ponens, modus tollens, denying the antecedent, affirming the consequent, and inconsistent forms of conditional arguments. Results demonstrated a significant advantage of the group over the container wording for some of the arguments. These results suggest that content that is specifically container-related does not provide a consistent advantage over other types of content, contrary to some theories of embodied mathematics.

Introduction

Conditional sentences are statements about relationships between an antecedent proposition (p) and a consequent proposition (q) that we will represent generically as “If p , then q .” In tests of conditional reasoning, individuals are asked either to construct conclusions or to confirm the validity of arguments made from premises that include a conditional sentence. Thus, conditional reasoning requires two cognitive processes: a) comprehending the premises and b) making inferences based on the premises (Marcus & Rips, 1979).

Conditional arguments can be valid or invalid, depending on how the conditional is interpreted. On a material conditional interpretation, the truth of a conditional depends on the truth values of the antecedent and the consequent. The only combination of antecedent and consequent that makes a material conditional false is one in which the antecedent is true, but the consequent is not (p , not q). Alternatively, a material biconditional interpretation dictates that a conditional is false unless both p and q are true, or both are false (p , q ; not p , not q).

Conditional reasoning experiments explore four main kinds of logical arguments. These are *modus ponens* (If p , q ; p ; therefore, q), *modus tollens* (If p , q ; not q ; therefore, not p), *denial of the antecedent* (If p , q ; not p ; therefore, not q) and *affirmation of the consequent* (If p , q ; q ; therefore, p). Given a material conditional interpretation of the premise, only modus ponens and modus tollens are valid arguments. Given a material biconditional interpretation, all four of the arguments are valid.

Some of these argument types are more difficult than others for participants to identify as valid or invalid. For instance, modus tollens arguments are typically more difficult for participants to verify than are modus ponens arguments, possibly due to the difficulty negation creates for individuals in reasoning, or the reversed order of the second premise and conclusion.

There is also a great deal of evidence that the content of the individual propositions affects the difficulty of the argument types (see Evans, Newstead, & Byrne, 1993, for a review). Participants often use their prior knowledge about the subject matter of the conditional to impose restrictions on the relationship between the antecedent and the consequent. In the case of conditional arguments, such as modus ponens and tollens, much of this evidence concerns how people’s knowledge of sufficiency and necessity alters their decisions about which arguments are valid (e.g., Ahn & Graham, 1999; Cummins, Lubart, Alksnis, & Rist, 1991; Thompson, 1994).

For example, if the antecedent is perceived as sufficient for the consequent (or, as in Cummins et al., 1991, participants cannot readily think of disabling factors that could prevent the antecedent from bringing about the consequent), participants will identify modus ponens and tollens arguments as valid. However, if participants view the antecedent as both sufficient and necessary for the consequent (or participants cannot imagine other possible causes of the consequent), they will identify denial of the antecedent and affirmation of the consequent as valid as well.

The present experiment investigates a different source of content effects, one that is explicit in current theories of embodied mathematics. Lakoff and Nunez (2000) hold that people understand abstract concepts, such as most mathematical concepts, by tying them to real-world experience---in effect grounding them in concrete reality. In the case of conditional reasoning, Lakoff and Nunez suggest that individuals use what they call the Container Schema, an internal representation of an object that can hold another. According to this theory, individuals translate any conditional argument into a mental representation of containment. For modus ponens, Lakoff and Nunez present the following concrete representation: “Given two Container

schemas A and B and an object X, if A is in B and X is in A, then X is in B.” Individuals then use what they know of containment to reason about the argument.

Certain types of content should be able to facilitate reasoning if individuals truly use the Container Schema. If Lakoff and Nunez are correct, it should be easier for individuals to respond to conditional statements that involve literal containers, such as buildings or boxes, than those that can only be interpreted as metaphorical or abstract containers, such as sets or groups.

Knowledge of containment should also lead to a material conditional interpretation in reasoning about conditionals involving literal containers. While being inside a room that is inside a building is sufficient for being inside that building, it is not necessary for it; one could just as easily be in another room in that same building. To the extent that the Container Schema governs more abstract content, people should also use the material conditional interpretation with groups or sets. While being a member of the Senate means that one is a member of the US government, being a member of the US government does not imply that one is in the Senate.

The present study investigates the effects of literal and abstract container-related content on accuracy in responding to conditional arguments. Unlike many earlier studies of conditional reasoning, this study employs conditional statements that express arbitrary relationships between the antecedent and the consequent, such that participants cannot import outside knowledge regarding the truth of the conditional. However, the antecedent and the consequent in each statement is always either a literal container (building) or an abstract container (a group). This allows participants to use all knowledge they possess of containment (e.g., a person cannot be both outside and inside a building) in order to see if such information allows participants to be more accurate in identifying the validity of logical arguments.

The Container theory predicts that participants should be more accurate in responding to conditional arguments when the antecedent and the consequent are literal containers than when they were both abstract containers. Lakoff and Nunez assert that “spatial logic is primary, and the abstract logic of categories is secondarily derived from it via conceptual metaphor” (Lakoff & Nunez, 2000, p. 45). If this claim is true, statements involving a relationship between two groups should be more difficult to reason about than statements involving a spatial relationship between two containers. We will evaluate accuracy in this case using the material conditional interpretation, since the Container Schema seems to presuppose this standard. We also expect to find that participants would be more accurate in responding to items where both the antecedent and the consequent were literal containers than when one was a literal container and the other an abstract container. The latter condition would require the combination of two different levels of abstraction into one coherent mental representation.

Methods

Materials

Instructions We told participants that they would see a series of three sentences and that they would have to determine if the third sentence (marked with a “C” and highlighted red) must be true whenever the first two are true. They were instructed to press the “y” key when the third sentence had to be true whenever the first two sentences were true and “n” when it did not. They were asked to move through the items as quickly and accurately as possible. They saw two example arguments, one where the third sentence did not logically follow and one where it did. These examples used logical forms unrelated to the conditional arguments (Either p or q; p; therefore not q).

Stimuli All participants received 120 arguments, each including a conditional statement, a minor premise, and a conclusion. There were 36 arguments of each content type: Container only (C), Group only (G) and Group by Container (G/C), plus 12 filler arguments of unrelated content. The container arguments were about two hypothetical spatial locations: the Hannley building and the Science Center. The group arguments were always about two hypothetical groups: the Spanish club and the Poetry club. The Group/Container arguments combined these locations and clubs (If the individual is in the Poetry club, the individual is in the Hannley building). The Group/Container condition was included as a control condition with the goal of allowing phrase content to differ as little as possible between the control and the Group and Container conditions.

Of the 36 arguments within each content type, there were three different types of phrasing for the conditional sentence: “If p, then q,” “All p are q,” “For all x, if x is p, then x is q.” Examples of the different content sets for each type of phrasing appear in Appendix A.

Within each set of twelve arguments for a given type of phrasing and content, four were inconsistent (INC) arguments (If p then q, p, therefore not q; If p then q, not p, therefore q; If p then q, q, therefore not p; If p then q, not q, therefore p), three were modus ponens arguments, three were modus tollens arguments, one was a denial of the antecedent argument, and one was an affirmation of the consequent argument. Thus, for each type of content, participants had the opportunity to respond to twelve INC arguments, nine modus tollens arguments, nine modus ponens arguments, three denial of the antecedent arguments, and three affirmation of the consequent arguments. We used this array of argument types to insure that there would be an equal number of valid and invalid arguments, according to the material conditional interpretation.

Participants received one of two versions of the entire stimulus set. The versions differed only in the way propositions were assigned to logical positions in the arguments. For example, one version assigned the proposition about being in the Spanish club to the

antecedent p and assigned the proposition about being in the Poetry club to the consequent q in the modus ponens argument “If p , then q ; p ; q .” The second version reversed these assignments.

Procedure

Participants received the instructions for the task and responded to the test items via computer. The participants read five screens of instructions, moving through them with the spacebar. During the task, participants studied each argument and pressed either the “y” or “n” key to indicate whether the third sentence was necessarily true whenever the first two sentences were true. As soon as the participant responded, the next argument appeared on the screen. Participants were not given any feedback on their performance. At the end of the 120 arguments, participants were presented with a “Thank you” screen and were then read debriefing information by the experimenter.

Participants

Participants were 30 Northwestern undergraduates who participated to fulfill a course requirement. Two additional participants were excluded due to computer malfunction.

Results

We calculated the proportion of correct responses as a function of content, argument type and phrasing. Contrary to the predictions of the Container theory, participants responded correctly to more Group items than to the Container or G/C items ($F(2, 56) = 3.15, p = .05$ ---see Figure 1). There was also a significant effect of type of logical argument on accuracy ($F(4, 112) = 31.31, p < .01$). The mean proportion of correct responses to denial of the antecedent and affirmation of the consequent across all participants was less than .50, while for modus ponens, modus tollens, and INC items, mean proportion correct across participants was over .80. These differences are generally consistent with those found in earlier studies (Evans et al., 1993). There was also a significant interaction of content and argument type ($F(8, 224) = 2.45, p = .015$). We will describe this interaction in connection with planned comparisons. There was no significant effect of phrasing on accuracy ($F(2, 56) < 1$), and this variable did not interact with content or type of logical argument. All the following accuracy analyses collapsed across phrasing.

Planned comparisons

More important, paired samples t-tests between types of content revealed significant differences. Participants responded correctly to significantly more denial of the antecedent items when they had Group content than when they had Container content (52.2 % vs. 36.7%, $t(29) = -2.83, p < .01$). Participants also responded correctly to significantly more modus tollens items for the Container content than for the G/C content (88.6% vs. 80.4% correct, $t(29) = 2.74, p = .01$). There was also a marginally

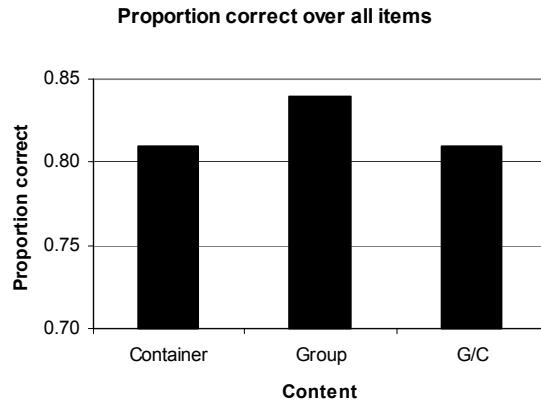


Figure 1: Proportion correct over all items by content

significant tendency to respond correctly to more modus tollens items for the Group content than for the G/C content (87.17% vs. 80.43%, $t(29) = 1.86, p = .07$). Figures 2 and 3 show the differences in proportion correct across content conditions for these arguments. There were no further differences due to content for any of the other argument types. Table 1 shows the percentage correct for each type of argument in each content condition.

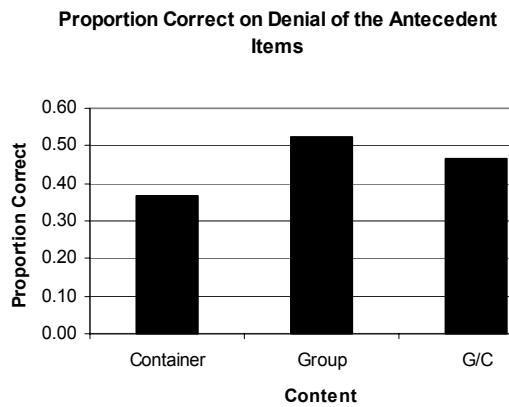


Figure 2: Proportion correct by content on the Denial of the Antecedent

Differences in interpretation

It is possible that differences in accuracy between content conditions arose from differences in the interpretation of conditionals based on their content. Within each content and phrasing group of twelve conditionals, participants were coded as having a material conditional interpretation if they answered “n” to all denial of the antecedent and affirmation of the consequent items, “n” to at least 3/4 of the INC items, and “y” to at least 2/3 of the modus ponens and tollens items. Participants were coded as having a material biconditional interpretation if they answered “y” to all denial of the antecedent and affirmation of the consequent

Proportion Correct on Modus Tollens Items

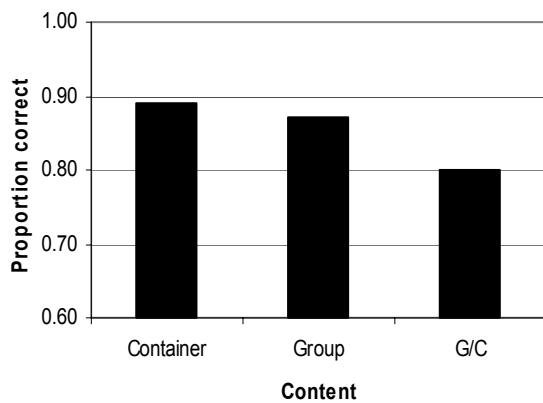


Figure 3: Proportion correct by content on Modus Tollens items

statements and to at least 2/3 of all modus ponens and tollens statements, and answered “n” to at least 3/4 of all INC statements. Participants with other patterns of responses were coded as having an “other” interpretation. Table 2 displays the percentage of the three types of interpretations for each content type (collapsed across phrasings).

A repeated-measures ANOVA with content and phrasing as repeated measures revealed no significant effect of either content or phrasing on the interpretations adopted by the participants, and no significant interaction of content and phrasing. However, despite the non-significant effect of content on interpretation, there was a notable tendency for more participants to adopt a material biconditional interpretation when evaluating the Container-related items than when evaluating the Group or Group/Container-related items. It is possible that this tendency led participants to be less accurate on the DA and AC items in the Container condition. One possible reason for this is that participants may have interpreted “If the individual is in the Hannley building, then the individual is in the Science Center” to

Table 1: Percentage correct for argument by content

Type of content	Type of Argument				
	AC	DA	INC	MP	MT
Container	41.1%	36.7%	88.1%	93.8%	88.6%
Group	55.6%	52.2%	86.4%	96.7%	87.2%
G/C	50.0%	46.7%	87.3%	95.9%	80.4%

AC = Affirming the Consequent, DA = Denying the Antecedent, INC = Inconsistent, MP = Modus Ponens, MT = Modus Tollens

Table 2: Percentage of material biconditional, material conditional and “other” interpretations by content

Interpretation Type	Content		
	Container	Group	Group/Container
MB	38.89%	24.44%	32.22%
MC	23.33%	31.11%	27.78%
Other	37.78%	44.44%	40.00%

mean that the Hannley building and the Science Center were the same building. To investigate this possibility, we have conducted a follow-up study using containers that are clearly distinct (a cardboard box and a wooden shipping crate). Although this change narrowed the difference between the Container and Groups conditions, there was no hint that containers promoted greater accuracy than groups.

Discussion

Although Lakoff and Nunez’s (2000) cognitive theory of mathematics would appear to predict an advantage for arguments that follow a Container Schema, the literal container items did not produce the highest accuracy levels in our experiment. Rather, participants showed best performance on items with Group content. The advantage for group content was most clear-cut for denial of the antecedent arguments. For modus tollens arguments, items with Container content produced slightly more correct responses, but even here there was no significant difference between Container and Group content items. Clearly, the literal container content did not produce a consistent advantage in accuracy over the other types of content. These results are surprising, given the suppositions of the Container Schema theory. It seems reasonable that a literal container would enable participants to form container schemas more efficiently and, thus, reason more accurately. In our study, this was not consistently the case.

The benefit of the Group over Container content is difficult to reconcile with the Container Schema idea. If we consider a group to be an abstract or metaphorical container, we would expect that processing arguments about groups would require additional steps to generate the metaphorical extension from the basic Container Schema. Our findings imply instead that there may be a processing advantage for conditional reasoning with *group* content. Further studies are planned to explore this possibility.

In this experiment, the effect of content appeared only for certain types of arguments. One reason for this is a ceiling effect for some of the remaining types. Responses to modus ponens items were correct 95% of the time across all content conditions. For INC items, the level of correct responses was also high: 87% across all content conditions.

Any effect of content would be expected to appear for the “harder” argument types: modus tollens, denial of the antecedent, and affirmation of the consequent. Modus tollens and denial of the antecedent items did show such an

effect. The reason for the absence of the effect on affirmation of the consequent items is unclear to us. We note, however, that both modus tollens and denial of the antecedent arguments involve two negatives (one in the minor premise and one in the conclusion). Affirmation of the consequent arguments, like modus ponens, contained no negatives. It seems possible that content in this study facilitated conditional reasoning by reducing the burden of processing negatives. It should be possible to investigate this effect in further work by introducing negatives in the antecedent and consequent of the conditional, as in some earlier studies of conditional syllogisms.

All of arguments in this study were container-related, on either a literal or metaphorical level. A potential objection to the evidence we have collected so far is that the Container Schema is so pervasive in people's thinking that any contact with the schema will improve reasoning. This objection may be difficult to handle, since remote extensions of the schema might always be possible. We note, however, that our results show reversals in performance (especially for denial of the antecedent arguments) that are unexplained on the hypothesis that any relation to containers boosts performance.

Conclusion

The current study cannot rule out the possibility that the Container Schema is sometimes helpful to people when they reason about conditionals. The findings do suggest, however, that Container Schemas are not the only source of content effects. The theory of Container Schemas suggests no reason for an advantage of abstract container-related content over literal container-related content and no reason why such effects should appear with some argument types but not others.

Acknowledgments

The authors would like to thank Jennifer Asmuth, Dan Bartels, Serge Blok, George Newman, Andrea Proctor and Eyal Sagi for their helpful suggestions and comments.

References

Ahn, W., & Graham, L. (1999). The impact of necessity and sufficiency in the Wason four-card selection task. *Psychological Science*, 10(3), 237-242.

Cummins, D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory and Cognition*, 19(3), 274-282.

Evans, J. St.B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human Reasoning*. Hillsdale, NJ: Erlbaum.

Lakoff, G., & Nunez, R. (2000). *Where Mathematics Comes From*. New York: Basic Books.

Marcus, S., & Rips, L. (1979). Conditional reasoning. *Journal of Verbal Learning and Verbal Behavior*, 18, 199-223.

Thompson, V. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition*, 22(6), 742-758.

Appendix A

Container content, "If" phrasing

If the individual is in the Hannley building, then the individual is in the Science Center.

Group content, "All" phrasing

All individuals in the Spanish club are in the Poetry club.

G/C content, "All ... if" phrasing

For all individuals, if they are in the Hannley building, then they are in the Poetry club.

***Note: All contents were presented in all phrasings**