University Libraries
University of Nevada, Las Vegas

University Libraries Lance and Elena Calvert Undergraduate Research Awards

Library Events

2011

The Effect of response format on syllogistic reasoning

Adam S. Billman
University of Las Vegas, Nevada, adambillman@hotmail.com

Follow this and additional works at: http://digitalscholarship.unlv.edu/award

Part of the Cognitive Psychology Commons

Repository Citation
Billman, A. S. (2011). The Effect of response format on syllogistic reasoning.
Available at: http://digitalscholarship.unlv.edu/award/6

This Research Paper is brought to you for free and open access by the Library Events at Digital Scholarship@UNLV. It has been accepted for inclusion in University Libraries Lance and Elena Calvert Undergraduate Research Awards by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.

The Effect of Response Format on Syllogistic Reasoning

Adam S. Billman

University of Nevada, Las Vegas

Author Note

I would like to thank Leslie Chavez, Hailey Connor, Dmitriy Kazakov, Jason Kelsey, Ashley Moreno, Nevenka Olivari and Jennifer Olsen for their assistance in collecting the data. I would like to thank Nicole Bies-Hernandez and Kris Gunawan for their advice on paper writing and statistics interpretation. I would especially like to thank David Copeland for his helpful guidance throughout all stages of this project. This project was not supported by any external funding sources.

All correspondence concerning this article should be addressed to Adam S. Billman, Department of Psychology, University of Nevada, Las Vegas, 4505 Maryland Pkwy Box 5030, Las Vegas, NV 89154. Email: billman3@unlv.nevada.edu

Abstract

This study examined the effect of response format on the conclusions drawn in a categorical reasoning task. In Experiment 1, participants were required to respond in either a multiple-choice format or were required to produce their own conclusion (without any choices presented). In the multiple-choice task, people responded with a wider variety of conclusions, more No Valid Conclusion responses, and had a smaller forward bias. In terms of theoretical predictions, Mental Model Theory was better supported in the multiple-choice condition and more conclusions corresponded with initial models in the production task while more conclusions corresponded with alternative models in the multiple-choice condition. The predictions made by the Probabilistic Heuristic Model were not affected by the conditions; however more min conclusions were found in the production task than the multiple-choice condition, and there was an increase in the p-entailment heuristic for the production task. Experiment 2, which included a persistent reminder about No Valid Conclusions in the production task, showed that No Valid Conclusions increased greatly for the production task and there was no longer a difference in forward bias. Also, neither theory showed any cross condition differences now.

The Effect of Response Format on Syllogistic Reasoning

Of critical importance to people's ability to successfully navigate the modern world is the capacity for both logical and abstract reasoning skills. Without these skills, it would be very difficult to generalize and carry information from one task to another. This would make learning difficult and time consuming. One of the types of reasoning that people use on a daily basis is categorical reasoning, which is used to find links between known information in order to derive connections with different pieces of information. The goal of this study was to discover what mechanism is likely at work when people perform categorical reasoning. In particular, this paper looked at changes in the kinds of conclusions that are drawn when an individual is confronted with different types of response formats. The conclusions drawn from these syllogisms will, in theory, reflect the kind of reasoning used and whether or not the type of reasoning changes with the response format.

Categorical Syllogisms

Categorical reasoning is a type of deductive logic that typically consists of two premises that make claims about various groups and their relations with one another. In a syllogism, these groups are typically represented by terms (e.g., A, B, or C), and quantifiers (i.e., All, Some, No, or Some… not) that are used to describe the relation these groups have with one another in statements called premises. For example, a pair of premises may state that 'All B's are A's' and that 'Some B's are C's'. These premises can be combined to form a syllogism that can be used to express a deductive argument. In such an argument, there will be two terms that appear once in the entire argument (i.e., the end terms: A and C) and one term that appears twice (i.e., the connecting term: B). Below is an example of a syllogism presented in a standard format:

All B's are A's

Some B's are C's

Therefore?

All reasoning performed on categorical syllogisms is deductive. It is important that people assume that both of the premises are true when trying to determine a conclusion, otherwise the premises would be violated. This means that a conclusion must be necessarily true given the premises that are presented in a syllogism. From the above premises, one can deduce that 'Some A's are C's'. To illustrate, if every B is an A, and if some of those B's (which are A's) are C's, then it follows that at least some A's must be C's. However, it should be noted that it is possible that people could respond that 'Some C's are A's'. The end terms can appear in either order, as long as the conclusion still holds true (this is discussed further below).

Categorical syllogisms can also come in several possible forms, and some have no deductively valid conclusion. Here is an example that is similar to the previous syllogism but in a different form:

All A's are B's

All C's are B's

Therefore?

Notice that the terms in the premises are both in different orders and the quantifiers have also changed. Unlike the previous syllogism, there is No Valid Conclusion for this syllogism. This is because unlike the previous syllogism, there are no answers that must be correct regardless of the possibilities. For example, it is possible that in addition to all A's being B's, all A's are also C's. It is also possible though that while A's and C's are B's, they are still mutually exclusive of each other (for example, chairs and tables are both furniture, but no chair is also a table).

This shows that there are several different orderings or syllogistic figures possible for the same syllogism. There are four syllogistic figures that premises can be placed in. These are AB-BC, AB-CB, BA-BC, BA-CB, and each represents a figure (i.e., ordering of terms in the two premises) that each premise could take. For example, the first syllogism above would be syllogistic figure BA-BC because both premises describe the linking term (B) at the beginning. The second syllogism would be syllogistic figure AB-CB because the first premise has the end term (A) in front and the second premise has the end term (C) first.

Finally, conclusions can be phrased with the terms presented in either order. For example, someone could derive a conclusion of ‘All A’s are C’s’ or ‘All C’s are A’s’. While these are not equivalent statements, the point is that it is possible for a conclusion to be written either in a forward or a reverse order. For example, in the following syllogism:

All A’s are B’s

All B’s are C’s

Therefore?

the conclusion can be framed as being either ‘All A’s are C’s’ (which is a valid conclusion) or ‘All C’s are A’s’ (which is not a valid conclusion). The first is in a forward order because the terms appear in the conclusion in the same order they appear in the premises (i.e., A-C); the second conclusion is in a reverse order because the terms are in the opposite order as they appear in the premises (i.e., C-A). Different orderings can affect a conclusion’s validity in some syllogisms, as in this example. However, in the first example above, order did not matter.

As stated earlier, in most research involving categorical reasoning, there are four possible quantifiers (All, Some, No, and Some… Not) that can be used in the two premises with each individual premise using one quantifier. When the four quantifiers are combined with the four

possible syllogistic figures, there are 64 possible combinations of syllogistic figures and quantifiers. Out of the 64 possible syllogisms, only 27 have valid conclusions with the rest having No Valid Conclusion.

Methods of Testing Syllogistic Reasoning

Similar to how there are different ways to design a college exam, there are a number of different ways that syllogistic reasoning can be tested. One way is to present the premises of a syllogism and to ask people to produce their own conclusion. An example of this is a set of experiments conducted by Johnson-Laird and Bara (1984) in which participants were asked to produce a conclusion from a given set of premises. In this experiment, participants were presented with cards that had single premises written upon them. After both of the premises were presented, the individuals then drew conclusions from scratch and wrote them down upon a piece of paper. Similar methods were also used by Goodwin and Johnson-Laird (2008) using transitive and pseudo-transitive inferences instead of straight forward categorical syllogisms (e.g., A is a blood relative of B, or A overtook B) and Jahn, Knauff, and Johnson-Laird (2007) which looked at how models were used in reasoning with spatial layouts.

One of the benefits of using a production task is a significant change in relative difficulty, which can be used to test how a theory holds up under tougher conditions. This is because all conclusions must be generated by the participants themselves. This increases difficulty because a production task requires additional resources (e.g., Copeland & Radvansky, 2004). However, a drawback of using a production task like this is that it may increase the difficulty of some tasks to the point that it can create floor effects. As participants try to derive a conclusion and try to determine the best solution to a problem, the difficulty of a problem could likely make it harder to render any decision of merit. In this scenario, people may not be able to come up with a

conclusion on their own. It may also cause participants to respond in unorthodox ways that are not score-able. For example, an individual may give an illegal answer using a quantifier not used in the task (e.g., “most” or “few”) or draw a conclusion using the connecting term instead of the end terms (e.g., ‘All A’s are B’s’ or ‘Some C’s are not B’s’).

A second technique that can be used to present syllogisms is to present them in a multiple-choice format. In this manner, the premises are presented and followed by a set of choices. Participants are asked to select the conclusion they thought best fit a given syllogism from the list of all nine possible conclusions. This method was used by Copeland and Radvansky (2004) and Copeland (2006) using all nine conclusions in a computer based test, and earlier by Dickstein (1978) who limited the choices to the conclusions that were in a forward order (i.e., this study omitted the four conclusions that are in a reverse order) or a response of No Valid Conclusion. This method allows researchers to score the task more effectively and quickly than a production task as multiple-choice requires far less interpretation than a production task and can often be scored mechanically. It also helps to systematize most scientific tests by ensuring that there is no possibility of an illegal response, because people must respond with one of the presented choices. However, some drawbacks are that multiple-choice tests may promote guessing because it can be easier to simply guess than it is to use reasoning.

There has been little research done to see if there is a difference between response formats. Hardman and Payne (1995) tested the production, multiple-choice and a third type of response format called a verification task against each other; the latter task involves asking people to evaluate whether a particular conclusion is true. In their second experiment, they presented a subset of the total 64 possible syllogisms to participants and asked them to either verify a conclusion that was presented with the syllogism, select from a list of five possible

answers (as used in Dickstein's (1978) study), or to produce their own conclusion from scratch. They found that there were differences between response formats with the greatest difference being between verification and the two other tasks. There were a number of limitations with this study. First, this study only used 42 of the 64 possible syllogisms. Second, the multiple-choice format was limited to only five choices as opposed to the full nine. Finally, this research was conducted in 1995 and so it did not examine some of the more widely accepted theories of today.

To our knowledge no research has been done to see if different theories of reasoning could be supported depending upon which format is used. For example, it could be that the reason one study provides support for one theory while it does not support another may have little to do with actual explanatory power, but more to do with the way participants are asked to respond. As mentioned before, this change may affect the difficulty of the task and could affect the kinds of strategies that are used to solve them. If this is the case, it can be assumed that the two are not compatible with one another. Further, it is possible that Hardman and Payne's (1995) study might have unduly limited the options by restricting the number and type of syllogisms used. In order to fully vet a theory of reasoning, the full set of 64 syllogisms should be examined. Because these theoretical explanations are an important part of understanding syllogistic reasoning, the next section explains the two most prominent theories.

Mental Model Theory

According to Mental Model Theory, reasoning is based upon representations that are mentally constructed (Johnson-Laird & Bara, 1984; Johnson-Laird, 1999; Bara et. al. 2001). The first thing an individual does is to interpret the information from the premises and use that information to construct a mental representation of what was described. After that first step, the individual then interprets the information in the next premise, and then combines the new

information with the model that has already been constructed. This process can be made more difficult if the premises are in a syllogistic figure where the order of the terms is not conducive to constructing a straightforward mental model. For example, the syllogistic figure AB-BC is straightforward because people can construct a model with A's and B's using the first premise, and then can connect the C's to the existing B's when interpreting the second premise. For the syllogistic figure BA-CB, people may prefer to start model construction with the second premise then the first, because the B's would then be linked when considering them in that manner. Therefore, construction of the model may need to wait until one or both of the premises have been converted into a syllogistic figure that is easier to deal with. This can be taxing on memory resources (Capon et al., 2007; Copeland & Radvansky, 2004). Once the transformation is complete, the individual then combines the separate information from both premises to form a new model. This process of transformation and augmentation is continued until all the premises are included in a final model.

Upon finishing the creation of a mental model, people can now form a conclusion. This conclusion can then be tested by attempting to build alternate models of the same premises that contradict the conclusion. The difficulty of this process is primarily determined by the number of models that can be produced from the same premises, with more models increasing the difficulty. If no model is constructed that can contradict the conclusion, then the conclusion is accepted. However, if the conclusion is contradicted, people will attempt to find another conclusion that fits the premises or else decide that there is No Valid Conclusion. It should be noted that some studies have shown that people have a tendency to rely on only a single model; in other words, even though more models may be possible, many people may not try to construct alternative models (e.g., Newstead, Handley, & Buck, 1999).

There are several predictions that can be made based on this theory (Johnson-Laird & Bara, 1984). One prediction is that there will be a preference for conclusions that are in the forward order (a forward bias). This is because producing a forward biased conclusion is less taxing in general than producing a reversed conclusion due to the fact that people generally read the 'A' premise first before reading the 'C' premise. As the syllogism is read, people will tend to construct the premises as they are read and so the model becomes oriented in a form that is read A-C rather than C-A. In addition, due to the difficulty of building multiple models, it is very difficult to completely verify a conclusion. This makes it more likely that the individual will fail to think of all the possible mental models that can be created from a set of premises, which would make an individual less likely to correctly respond with an answer of No Valid Conclusion and answer with an answer that may not be valid.

Probabilistic Heuristic Model

Recently, there has been a movement to describe many mental operations as being a series of probabilistic judgments rather than based on mental models or formal logic rules (Chater & Oaksford, 1999). The Probabilistic Heuristic Model looks at reasoning as being more akin to the use of fast and frugal heuristics based on probabilities rather than the complex and difficult task of manipulating mental tokens as proposed by Mental Model Theory.

The first step an individual takes, according to this theory, is the generation of a conclusion using a set of heuristics to help create a conclusion that is probably valid (Chater & Oaksford, 1999). There are three heuristics that can be used, the first of which is the min-heuristic. This heuristic favors the quantifier that is the *least* informative. For example, according to this theory, the quantifier 'All' is more informative than the quantifier 'Some', and 'Some' is more informative than 'No' which is more informative than 'Some… Not'. The quantifier that

the first heuristic favors is called the min quantifier. For example, out of the two quantifiers ‘All’ and ‘Some’, ‘Some’ would be favored because it is less informative than the quantifier ‘All’. Once this quantifier is determined, then it becomes the preferred quantifier to be used in the conclusion. The second heuristic is the P-entailment heuristic. After determining the min quantifier, an individual will then try to figure out, using the least informative premise, what other conclusions will most likely also be true. For example, if the min heuristic leads to a conclusion that all A’s are C’s, then it is likely (P-entailed) that some A’s are C’s. Thus, this second heuristic can also be used to derive the quantifier that is used to create the conclusion. The last generation heuristic is the Attachment Heuristic. This heuristic is used to determine the order of the terms in the conclusion. This heuristic will attach the quantifier from the min premise (i.e., the premise with the min quantifier) to the term of the most informative premise unless the quantifier from the min premise has the last term (usually C) as its subject. For example, if the last premise of a syllogism is “Some C’s are not B’s” then the attachment heuristic will attach the min premise (Some…Not) onto the term C rather than the term A. However if that premise was instead “Some B’s are not C’s” then attachment will occur with the term A and not C.

There are two test heuristics that are employed by an individual to test if a conclusion that has been derived is likely to be true. The first test is called the max-heuristic. This test determines how confident an individual is in a particular conclusion. This depends on the difference in informativeness between the min and the max premise. For example, if the most informative premise was ‘All A’s are B’s’ and the conclusion used the quantifier ‘Some…Not’ then the individual who made that conclusion would not be very confident in that answer due to the large difference in informativeness between the most informative premise and the

conclusion. However if the conclusion used the quantifier ‘All’ instead, then the individual would be much more confident and would likely respond with that answer. The second test is the O-heuristic. This test seeks to avoid producing ‘Some… not’ conclusions due to their lack of informativeness. The lack of informativeness tends to make the individual uncomfortable with the conclusion. This means that when a conclusion is especially uninformative (as a ‘Some… Not’ conclusion would be), then the individual will want to reject the conclusion and respond with No Valid Conclusion.

As can be seen in the above description, the Probabilistic Heuristic Model differs from Mental Model Theory. According to Mental Model Theory, people use the quantifiers to represent the terms, and to then determine their relationships to one another. However, according to the Probabilistic Heuristic Model, people make probabilistic judgments about what is more likely based solely on the quantifiers that are presented. These differences lead Probabilistic Heuristic Theory to predict different conclusions than Mental Model Theory. The Probabilistic Heuristic Model can be used to make several predictions. Due to the attachment heuristic, an individual would be almost just as likely to produce a conclusion that is in a forward direction as in a reverse direction. Also, due to the probabilistic nature of the heuristics, there is more room for variation. This means that there would likely be a larger variety of answers that would likely be given in any group.

While there is a lot of research on categorical syllogisms (Chater & Oaksford 1999; Copeland 2006; Copeland & Radvansky 2004; Dickstein 1978; Johnson-Laird & Bara 1984), there has only been one study (Hardman & Payne, 1995) that made a direct comparison of the response format methods used in most research to see if they are indeed comparable to one another. However, as pointed out earlier, that study needs to be expanded and updated by using

the full set of syllogisms, the full set of responses, and by examining performance in the context of more recent theories. Without a comparison, there is no reason to assume that the conclusions drawn by one study will be either compatible or comparable to the conclusions of another study.

Experiment 1

The purpose of this study is to examine whether or not there are significant differences in answers that tend to be given between multiple choice and production task formats. In Experiment 1, this was tested by having different groups of undergraduate students answer a full set of 64 categorical syllogisms. The two groups were required to answer the question by indicating the correct answer from the list of answers (i.e., multiple-choice format) or by typing it in themselves without the inclusion of the list of possible answers (i.e., production format).

The two theories in question have led to the following hypotheses. First, these response formats will differ in the number of conclusions given per syllogism with multiple-choice having a wider variety of answers given. This is because multiple-choice will have all the possible answers displayed and this should encourage participants to consider different answers. The number of No Valid Conclusions should be higher for the multiple-choice condition than the production task for much the same reason as the last prediction. The answers being displayed on the screen in the multiple-choice condition should lessen the working memory load for this condition, allowing people to try different solutions. Along this line, because the production task taxes working memory more, fewer answers should be considered and there should be a larger forward bias.

With regard to the theories, it is predicted that the amount of support a theory has will depend largely upon the answering format. One prediction for Mental Model Theory is that participants will rely on the initial model more often in the production task than in the multiple-

choice condition, mainly because the former is more taxing on working memory. A second prediction is that additional models will be more likely to be considered for the multiple-choice condition than the production task. This is because the presence of the choices will increase the likelihood that participants will try to falsify their conclusions, and will reduce the difficulty because the participants no longer must store the information separately in working memory.

For the Probabilistic Heuristic Model, the number of conclusions corresponding to the min heuristic should be greater with the production task. Also the number of p-entailment conclusions should also increase with the production task relative to the multiple-choice task. This is because the participants had to both come up with and remember their conclusion during the reasoning process. This should create higher difficulty levels, which participants would respond to by using p-entailment. This is because a participant might go with a weaker conclusion that has a higher likelihood of being true than a stronger, more informative conclusion. Or, to state it more simply, people may be more likely to guess with a highly probable conclusion when the task is more difficult.

Method

Participants

Forty-eight undergraduate students (24 males, 24 females) from the University of Nevada, Las Vegas participated. All of them participated for partial credit in a psychology course. Ages ranged between 18 and 34 years with a mean of 20.9 years.

Materials and Procedure

The materials included all 64 possible categorical syllogisms. Each syllogism used the words ‘Hikers’ and ‘Readers’ as the end terms and the word ‘Swimmers’ as the connecting term. For example:

All Hikers are Swimmers

All Swimmers are Readers

Therefore…?

These hobby terms have been used in past studies (e.g., Copeland & Radvansky, 2004) to ensure that each premise and conclusion could be considered plausible.

Each participant used a standard computer for all testing and were instructed to determine conclusions using no additional materials (i.e., no paper or pencil was allowed). Also, the order of the presented syllogisms was randomized for each participant. Each participant signed the standard consent form and was then instructed on the basics of syllogisms. For example, they were told what terms are, what a quantifier is, and what a syllogism is. Then they were taught what a conclusion is, how to create one, what kinds of orders a conclusion could take, and what makes a conclusion valid. However, they were not given any information on the basics of solving categorical syllogisms (i.e., they were not provided any information about strategies) nor were there practice problems for the participants prior to testing.

After all the instructions were read, the participants began working on the problems, completing all of them in one sitting. Each participant was randomly placed into either the multiple-choice or the production task condition. In the multiple-choice condition, each participant was required to select the conclusion they thought best fit the premises in the syllogism from the list of the nine possible conclusions. For example:

1) All Hikers are Readers
2) All Readers are Hikers
3) Some Hikers are Readers
4) Some Readers are Hikers

5) No Hikers are Readers
6) No Readers are Hikers
7) Some Hikers are Not Readers
8) Some Readers are Not Hikers
9) No Valid Conclusion

In contrast, for the production task, participants were asked to type in the conclusion they thought was correct given the premises into a dialog box without the presence of the possible conclusions. Only data from participants that provided legal conclusions (i.e., conclusions using one of the 4 quantifiers and containing the two end terms) were scored. The answers were categorized as to the type of answer given (e.g. ‘All A’s are C’s’ or ‘No C’s are A’s’).

The data was collected and used to find several measurements. First we analyzed the total number of conclusions that were given for each syllogism. This was found by tallying the number of conclusions that received at least one response (i.e., from at least one participant) for each syllogism. We also looked at the number of No Valid Conclusions given per syllogism. Finally we computed the forward-bias by taking the number of responses that were in a forward order (A-C) and subtracting the number that were found to be of the reverse order (C-A). Accuracy was also measured as the number of correct responses (out of 64).

Results and Discussion

All analyses were conducted as item analyses (see Table 1), matching them for the two conditions (multiple-choice / production). To compare the variability of responses, the number of different conclusions given by participants was determined for each syllogism. For example, if for a particular syllogism, 21 participants responded with “All A are C”, 3 responded with “All C are A”, and no other conclusions were given by participants, then this syllogism would be scored

as a 2. If, on the other hand, at least 1 participant responded with each conclusion for a particular syllogism, then that syllogism would be scored as a 9 (because there are 9 possible conclusions). A paired samples t-test showed that people answered with a wider variety of conclusions for the multiple-choice (M = 5.94, SE = 0.15) than the production task (M = 5.33, SE = 0.19), $t(63) = 3.51$, $p < .01$; the reported means refer to the mean number of different conclusions per syllogism. People also were more likely to respond with No Valid Conclusion in the multiple-choice (M = 4.23, SE = 0.33) than the production task (M = 1.88, SE = 0.25), $t(63) = 9.95$, $p < .001$; the reported means refer to the mean number of people responding with No Valid Conclusion for each syllogism. There was also a higher proportion of A-C conclusions given relative to C-A conclusions for the production (M = 8.56, SE = 0.94) than the multiple-choice task (M = 2.48, SE = 0.96), $t(63) = 9.38$, $p < .001$; the reported means refer to the mean difference score for each syllogism, subtracting the number of C-A conclusions from the number of A-C conclusions. Finally, people in the multiple-choice condition (M = 33.01%, SE = 2.99) were more accurate than those in the production condition (M = 24.74%, SE = 3.52), $t(63) = 7.21$, $p < .001$.

Because the goal of the current study was not to distinguish whether one theory may be better than the other, comparisons of the theories were not conducted (see Copeland & Radvansky, 2004, and Copeland, 2006, for direct comparisons of these theories). Instead, we focused on comparing the performance of a particular theory across the different response formats (see Table 2). The percentage of responses that corresponded to those predicted by Mental Model Theory was greater for the multiple-choice condition (M = 79.55%, SE = 1.22%) than for the production task (M = 74.59%, SE = 2.18%), $t(63) = 2.94$, $p < .01$. Participants were more likely to give an answer that corresponds to an initial model answer in the production task

(M = 60.00%, SE = 3.08%) than in the multiple choice task (M = 54.59%, SE = 2.79%), $t(63) = -2.97$, $p < .01$. Finally the multiple-choice condition had more answers corresponding to multiple models (M = 25%, SE = 2.29) than the production task (M = 14.72%, SE = 1.96%) $t(63) = 8.89$, $p < .001$.

The percentage of responses that corresponded to those predicted by the Probabilistic Heuristic Model did not differ between the multiple choice condition (M = 88.41%, SE = 1.16%) and the production task (M = 88.58%, SE = 1.5%), $t(63) = 0.18$, $p = .854$. However, the number of conclusions that corresponded to the min heuristic was greater for the production task (M = 60%, SE = 3.08%) than the multiple choice condition (M = 54.59%, SE = 2.79%) $t(63) = 2.97$, $p < .005$, and the p-entailment heuristic was greater for the production task (M = 20.84%, SE = 2.31%) than for the multiple-choice task (M = 16.19%, SE = 1.64%), $t(63) = 2.59$, $p < .05$.

From the results above, it seems that a major cause of the differences seems to be that people in the production task are failing to invalidate their initial models (from Mental Model Theory) or are seeking weaker conclusions using p-entailment (from Probabilistic Heuristic model). From the standpoint of Mental Model Theory, this result is to be expected as the added work necessary in generating additional models while still remembering the conclusion one is trying to test is very difficult based on working memory limitations. So it seems that because of this difficulty participants are not trying to generate models beyond the first one. This effect is lessened in the multiple choice condition because the participant is no longer required to generate their own conclusions and furthermore, the constant presence of these choices encourages participants to try to falsify their conclusions more in order to come up with the ‘better answer’.

Probabilistic Heuristic Model does not predict a difference in difficulty due only to the presence (or lack of) a list of choices. However, there is an increase in the usage of p-entailed

answers, which suggests that participants are weakening their conclusions in the production task. An explanation of this effect is that because the participants had to both come up with and remember their conclusion during the reasoning process, this created unusual difficulty levels that participants responded to by using p-entailment. This is because a participant might go with a weaker conclusion that has a higher likelihood of being true than a stronger, more informative conclusion.

As the results from Experiment 1 show, the two response formats are not equivalent in either the descriptive data or the theory predictions. This means that any results from one experiment are not equivalent to another unless both experiments use the same type of response format. These results could explain the differences seen between some of the studies that can be found in the literature and perhaps explain why sometimes a theory is supported in one study and another theory in a different study.

Experiment 2

Of course, the fact that a difference exists does not explain why it exists. In Experiment 1, it was suggested that the reason these results were different between the two methods of testing is because the participants are either forgetting or not attending to all the possible answers. For example, in the syllogism:

Some A’s are B’s

Some B’s are not C’s

Therefore….?

Someone should conclude that there is No Valid Conclusion. However, in the production task, because no choices are presented, it is possible that the reasoner may simply have forgotten that this response was a possibility and hence did not respond properly. In order to discover if this

was a possible cause of our results, we created Experiment 2. The goal of Experiment 2 was to examine this possibility by adding a persistent reminder regarding the possibility that some syllogisms may have No Valid Conclusion. It was predicted that this change would increase the likelihood that, according to Mental Model Theory, participants in the production task would try to falsify their initial models and consider alternative models. For the Probabilistic Heuristic Model, participants should consider additional heuristics beyond the min heuristic and p-entailments than they did in Experiment 1.

Method

Participants

Forty-eight undergraduate students (17 males, 31 females) from the University of Nevada, Las Vegas participated. All of them participated for partial credit in a psychology course. Ages ranged between 18 and 40 years with a mean of 20.5 years.

Materials and Procedure

The materials used and the procedure was identical to Experiment 1 except that in Experiment 2, the production task included a persistent reminder that No Valid Conclusions were a possible response. This reminder was presented on the screen with every syllogism. The multiple-choice condition was identical to Experiment 1.

Results and Discussion

The analyses for Experiment 2 (see Table 3) were similar to those conducted in Experiment 1. In Experiment 2, we found that the variability was still significantly higher in the multiple choice condition (M = 5.62, SE = .16) than the production task (M = 5.06, SE = .16), $t(63) = 3.02$, $p < .01$. As predicted, the number of No Valid Conclusions increased in the new production task with the production task (M = 4.23, SE = .40) now being greater than the

multiple choice condition (M = 3.43, SE = .38), $t(63)$ = 3.44, $p < .01$. There was no longer a difference between the amount of forward bias for the multiple choice (M = 5.25, SE = 1.11) and production task (M = 5.98, SE = 1.09), $t(63)$ = 1.14, p = .26, nor was there a difference in accuracy between the multiple choice condition (M = 32.23%, SE = 3.33%) and the production task (M = 32.94%, SE = 3.34%), $t(63)$ = 0.47, p = .64.

The same comparisons from Experiment 1 were made for the predictions of Mental Model Theory and the Probabilistic Heuristics Model (see Table 4). There were no longer as many differences between the two conditions for either theory. For Mental Model Theory, there was no difference between the multiple choice condition (M = 78.73%, SE = 1.53%) and the production task (M = 76.98%, SE = 2.17%), $t(63)$ = 1.00, p = .32 regarding the percentage of responses that corresponded to its predictions. There was also no difference between the prevalence of initial model answers in the multiple choice condition (M = 55.80%, SE = .304%) and the production task (M = 53.77%, SE = 3.48%) $t(63)$ = 1.15, p = .26, nor the usage of additional models between multiple choice (M = 23.03%, SE = 2.63%) and the production task (M = 23.19%, SE = 2.41%) $t(63)$ = 0.11, p = .91.

For the Probabilistic Heuristics Model, overall there was no difference between the multiple choice condition (M = 88.53%, SE = 1.24%) and the production task (M = 90.55%, SE = 1.39%) $t(63)$ = 1.36, p = .18. There was no difference between the usage of the min heuristic between the multiple choice condition (M = 55.8%, SE = 3.03%) and the production task (M = 53.77%, SE = 3.47%) $t(63)$ = 1.14 p = .25 nor was there a difference in the p-entailment heuristic between the multiple choice condition (M = 18.47%, SE = 1.98%) and the production task (M = 19.02%, SE = 2.26%), $t(63)$ = 0.36, p = .72.

As the data from Experiment 2 showed, one of the major causes of the differences in Experiment 1 seems to be that participants were either forgetting or failing to attend to the possibility of a No Valid Conclusion when the production task was used. This is seen in the fact that there was a difference in both the number of No Valid Conclusions and also in the slight difference in the total number of conclusions drawn. Indeed, it seems that simply adding the reminder that there can be No Valid Conclusions had an effect on all measurements we used, though the differences in forward bias may simply be caused by an overall decrease in conclusions other than No Valid Conclusion.

General Discussion

The current study examined whether there was a difference between the production task and the multiple-choice conditions when participants performed categorical reasoning. We predicted that there would be differences between the two conditions in the types of conclusions drawn, and that these differences would show in the theories. The results showed a difference between the two conditions when looking at the conclusions drawn, particularly in regards to No Valid Conclusions and the forward bias. These differences disappeared when persistent reminders about No Valid Conclusion being a possible response were added, suggesting that memory or awareness effects were a possible primary cause. This means that the two response formats are not equivalent and so experiments with one format should not be directly compared with another experiment that used a different format. Consistent with this idea, the current study showed that these differences could actually affect the amount of support a theory can receive from the data. Fortunately, it seems that reminders may control the differences somewhat.

This study went beyond Hardman and Payne's study (1995) in several ways. First, in the previous study, Hardman and Payne sought to control the answers participants gave by limiting

the number of syllogisms to only those that would elicit a forward bias and restricting the list of conclusions in the multiple choice condition to those of a forward bias, whereas the current study did not make any restrictions and used the full 64 possible syllogisms and all nine possible conclusions. Second, the original study did not look into the effects these different response formats have upon any theories of reasoning. In contrast, the current study examined the effects of these response formats on two prominent theories of reasoning, the Probabilistic Heuristic Model and Mental Model Theory.

With regard to both theories of reasoning, it should be noted that differences in format may have actually changed the strategies people use. With reference to Mental Model Theory, there was a large increase in the use of additional models when participants were in the multiple-choice condition. This suggests that participants' working memory was being taxed, and in response participants were not able to be as thorough in their search for additional models. When all of the possibilities were presented in the multiple-choice condition, however, participants began to create additional models in order to try to falsify their prior conclusions. This difference in additional models was very clear in the current study which suggests that Mental Model Theory should seek to explain this effect in newer implementations. This is especially true because the level of support varied with the use of additional models, which suggests that this effect is very important in the reasoning process.

As far as the Probabilistic Heuristic Model is concerned, there was little change between conditions except that p-entailment usage was increased in the production task. This suggests that when the task is more difficult, participants might be weakening their conclusions for a conclusion that is more likely to be acceptable. Relative to Mental Model Theory, the effects of response format on the predictions of the Probabilistic Heuristic Model were far less pronounced.

This suggests that the theory may be a more complete theory that might be capable of explaining more response formats than Mental Model Theory. There were still differences in the usage of p-entailment though and this should be examined further in future studies.

Conclusion and Future Directions

These findings should be considered in future research in order to prevent confounds from reducing the validity of valuable research. This study showed that one cannot look at Johnson-laird and Bara's (1984) results, which used a production format, and make direct comparisons to the data from Chater and Oaksford's (1999) study, which used a multiple-choice format. However, this does not mean that only one response format should be used in all research though. As people perform reasoning in many different formats in real life, research should also include several types of formats. Furthermore, all theories of reasoning should be able to adequately explain the differences in answers given in all response formats. If participants answer a question differently in one format than in another, a good theory of reasoning should be able to explain why this is so. However, researchers should keep in mind that a study that uses one response format is not comparable to another.

Further research should be directed toward looking at what other factors may cause the differences between response formats. For example, it may be worth examining whether the persistent reminder used for No Valid Conclusions will have similar balancing effects when used to remind participants about forward bias. It may also be worth looking into the possibility that participants could be using different reasoning strategies depending upon the situation. It is suggested that additional research could be used to discover if that is a possibility.

References

Bara, B.G., Bucciarelli, M., & Lombardo, V. (2001). Model theory of deduction: A unified computational approach, *Cognitive Science*, *25*, 839-901.

Capon, A., Handley, S., & Dennis, I. (2003). Working memory and reasoning: An individual differences perspective, *Thinking and Reasoning*, 9 (3), 203–244

Chater, N., Tenenbaum, J., & Youille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Psychology, 10,* 287-291.

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, *38*, 191-258.

Copeland, D. E. (2006). Theories of categorical reasoning and extended syllogisms. *Thinking & Reasoning, 12*, 379- 412.

Copeland, D. E., & Radvansky, G.A. (2004). Working memory and syllogistic reasoning. *The Quarterly Journal of Experimental Psychology, 57A*, 1437-1457.

Dickstein, L. (1978). The effect of figure on syllogistic reasoning. *Memory & Cognition, 6*, 76-83.

Goodwin, G.P., & Johnson-Laird, P.N. (2008). Transitive and pseudo-transitive inferences. *Cognition, 108*, 320-352.

Hardman, D. K. & Payne, S. J. (1995). Problem difficulty and response format in syllogistic reasoning, *The Quarterly Journal of Experimental Psychology Section A*, *48* (4), 945-975.

Jahn, G., Knauff, M., & Johnson-Laird, P.N. (2007). Preferred mental models in reasoning about spatial relations. *Memory & Cognition, 35*, 2075-2087.

Johnson-Laird, P.N., & Bara, B. G. (1984). Syllogistic inference. *Cognition, 16*, 1-61.

Johnson-Laird, P. N. (1999). Deductive Reasoning. *Annual Reviews of Psychology, 50*, 109—35.

Newstead, S. E., Handley, S. J., & Buck, E. (1999). Falsifying mental models: Testing the predictions of theories of syllogistic reasoning. *Memory & Cognition, 27*, 344-354.

Table 1

The number of conclusions drawn per syllogism, No Valid Conclusions, amount of forward bias, and accuracy in Experiment 1 (Standard Error in parentheses)

	Multiple-Choice	Production
Number of Conclusions	5.93 (.15)	5.32 (.18)
No Valid Conclusions	4.23 (.33)	1.87 (.25)
Forward Bias	2.48 (.95)	8.56 (.93)
Accuracy	33.01% (2.99)	24.74% (3.52)

Table 2

Percentage of responses that correspond to predictions made by Mental Model Theory and the Probabilistic Heuristics Model in Experiment 1 (standard error in parentheses)

	Multiple-Choice	Production
Mental Model Theory (Overall)	79.55% (1.22)	74.59% (2.18)
Initial Model	54.59% (2.80)	60.00% (3.09)
Second or Third Model	25.00% (2.30)	14.72% (1.96)
Probabilistic Heuristics Model (Overall)	88.41% (1.16)	88.58% (1.50)
Min Heuristic	54.59% (2.80)	60.00% (3.09)
Probabilistic Entailment	16.19% (1.64)	20.84% (2.32)

Table 3

The number of conclusions drawn per syllogism, No Valid Conclusions, amount of forward bias, and accuracy in Experiment 2 (Standard Error in parentheses)

	Multiple-Choice	Production
Number of Conclusions	5.62 (.16)	5.06 (.16)
No Valid Conclusions	3.43 (.37)	4.32 (.40)
Forward Bias	5.25 (1.1)	5.98 (1.08)
Accuracy	32.23% (3.32)	32.94% (3.33)

Table 4

Percentage of responses that correspond to predictions made by Mental Model Theory and the Probabilistic Heuristics Model in Experiment 2 (standard error in parentheses)

	Multiple-Choice	Production
Mental Model Theory (Overall)	78.73% (1.53)	76.98% (2.17)
Initial Model	55.80% (3.04)	53.77% (3.48)
Second or Third Model	23.03% (2.63)	23.19% (2.41)
Probabilistic Heuristics Model (Overall)	88.53% (1.24)	90.55% (1.40)
Min Heuristic	55.80% (3.04)	53.77% (3.48)
Probabilistic Entailment	18.47% (1.98)	19.02% (2.26)